1970

# Linear regression models of sound velocity in the North Atlantic Ocean below a critical depth

Richard Roland Kunkel

146

# LINEAR REGRESSION MODELS OF SOUND VELOCITY IN THE NORTH ATLANTIC OCEAN BELOW A CRITICAL DEPTH

by

RICHARD ROLAND KUNKEL, 1945-

A DISSERTATION

Presented to the Faculty of the Graduate School of the

UNIVERSITY OF MISSOURI - ROLLA

In Partial Fulfillment of the Requirements for the Degree

DOCTOR OF PHILOSOPHY

in

GEOPHYSICAL ENGINEERING



1970

Hughes M Zenor
Advisor

S. A. [illegible]

A K Rigler

Gerald B. Rupert

C. Y. Ho

Lee J Bain

# ABSTRACT

Sound velocity models currently in use by the U. S. Navy require difficult data sampling at great expense. This study is to investigate the feasibility of a sound velocity model dependent only upon position that would be valid for a selected, reasonably large, near homogenous area of the ocean. The most important position parameters are latitude, longitude, and depth.

To accomplish this objective a critical depth of 2500 meters was established for an $8^{\circ}$ by $20^{\circ}$ rectangle of the Sargasso Sea which lies within the North Atlantic Ocean. The critical depth is defined as the depth at which the characteristics of the area of the ocean being considered become stable enough to predict sound velocity as a function of position within a specified accuracy.

Such a model was produced with the aid of new concepts of independent term generation within stepwise multiple regression software that allowed a study of a variety of models with unprecedented ease.

## ACKNOWLEDGEMENTS

The author wishes to express his sincere appreciation to Dr. Hughes M. Zenor for his suggestions and encouragement throughout this study.

Appreciation is extended to Mr. Richard A. Bolton of the United States Naval Oceanographic Office who originally suggested the undertaking of this study, facilitated the obtainment of data, supplied graphic software, and conceived of many new ideas made available in Appendix A.

Cordial recognition is due Dr. A. K. Rigler of the Computer Science Department for his continual daily assistance in analyzing and perfecting this study.

Finally, deepest gratitude is expressed to the author's wife, Karen, and son, Timothy, for their patience and understanding during these years of study, and to the writer's wife special recognition is deserved for the many hours she spent at his side during the construction and final typing of this paper.

# TABLE OF CONTENTS

# LIST OF ILLUSTRATIONS

# LIST OF TABLES

## LIST OF MAPS

# I. INTRODUCTION

Although oceanography, the application of sciences to the phenomena of the oceans, is an old science in that man wondered about the oceans of the world long ago; it is also a very new science. Until recently only the oceans' surfaces had been explored to any extent. Men knew little of the depths, little of the details of these great water filled basins that cover 70.8 percent of the surface of the earth.

Acquiring adequate knowledge of the subsurface of the oceans is a task to which the United States Navy has committed itself. Within the scope of this task is the determination of the speed of sound at any given depth in the oceans of the world. One recognizes the importance of this problem when reviewing the statistics of submarine warfare during World War II. However, the problem of determining sound velocity in the oceans has become more pertinent and important in a peacetime atmosphere with man's insistance on exploring the oceans and their boundaries.

Since the oceans of the world cover over two-thirds of the earth's surface, the investigation of sound velocity in all oceans can not be undertaken under one project. Furthermore, the composition of the oceans are so variable that even one body of water is usually subdivided into specified sections. Thus, one must look at only a certain section of a body of water at a time in order to obtain an accurate sound velocity model for that section. The number of sections needed seems to be dependent on the variability of the body of water being considered.

The Atlantic Ocean alone covers 16. 2 percent of the earth's surface. Because of its dimensions, it is divided into the North and South Atlantic. Again the North Atlantic's size warrants that it be subdivided and that only one subdivision be investigated at a time. It is the purpose of this study to investigate sound velocity in a section of the North Atlantic Ocean, and in turn to determine a more reasonable and feasible sound velocity model than those already in existence for the section considered.

In order to determine whether a model is more reasonable or feasible some criteria must be established for this end. The criteria used in this paper is the number of independent terms necessary to implement the model, the expense in acquiring data for the model, and the accuracy of the model. All of the above have been obstacles to the Navy in utilizing the sound velocity models already in existence.

In the past the Navy has developed analytical, graphical, and numerical sound velocity models for sea water. Much work has been done with all three types of models and the comparisons among them. In this paper numerical methods will be employed through using stepwise multiple regression analysis. However, new numerical techniques in term generation within the stepwise multiple regression have been created in order to help eliminate the trial and error methods of deciding which terms should appear in the regression model and facilitate the laborious procedures of data preparation.

Since underwater sound velocity is a function of many independent variables, its method of obtainment still puzzles the minds of men today. Nevertheless, if one could definitely conclude that only certain independent variables were necessary for the determination of sound velocity in sea water, the problem would still not be solved. Many independent variables that are now considered essential to compute sound velocity such as salinity, temperature, and pressure can be obtained by only crude methods of data collection at great expense. Data collection becomes more expensive as the depth of the observation increases. No reliable method exists for the computation of sound velocity at a given position that eliminates the need for on-location measurements of the desired independent variables. Thus, expensive data collection and data storage must precede deep underwater calculation of sound velocity.

If one would investigate sound velocity models already developed, complex equations with many independent variables would be encountered. Furthermore, each model would be valid for only a small section of the body of water being considered. Since the composition of sea water is so variable, it limits a given model from being accurate over a large section. Thus, a number of models must be incorporated in order to determine sound velocity for a sizeable body of water.

A further investigation of known models of sound velocity and graphs of the important independent sound velocity variables would illustrate a more predictable or stable system as the depth increases for a particular latitude and longitude of the body of water in question. This phenomena indicates that at near surface depths the role of the independent variables are quite important for it is at these shallow depths that these variables are most erratic. Much of this can be accounted for by the seasonal change of the earth's atmosphere.

Therefore, it seems reasonable that some critical depth should exist such that below this depth it becomes feasible to build a simplified model of sound velocity for large portions of the ocean. By large portions is meant portions larger than the sections used for sound velocity models which are accurate for all depths. Of course there are exceptions to this general observation. For example, whenever a discontinuity exists such as at the interface of two different bodies of water, the above may not apply.

Thus, the author of this paper believes that this critical depth should be established; that the more accurate and well-known equations for the calculation of sound velocity should continue to be used above this critical depth where data collection is less expensive; and that a new model should be employed below this critical depth which would contain fewer independent variables. It is also below this critical depth that the collection of data is most expensive. If this new model is to be a positive contribution, then it must be accurate, hold for large portions of the body of water being considered, and be less expensive to implement.

The most expensive part of data collection, along with all of its present difficulties, could be discarded if this model could completely eliminate the need for on-location measurements. This would definitely indicate a model dependent only upon position. The more important position parameters for this study are latitude, longitude, and depth. The feasibility of this is discussed in Chapter II and the model itself is investigated in Chapter III. Even if only a few independent variables could be dropped, substantial savings would be realized.

The massiveness of the oceans creates an obstacle when handling data for oceanographic problems. While the data for a certain section of the ocean might be relatively sparse, the total amount of data for the section is extremely voluminous. This plus other factors such as limited computer facilities, have defined the size of the section of the North Atlantic for this study which could reasonably be examined in detail. A $8^{\circ}$ by $20^{\circ}$ rectangle off the east coast of the United States has been chosen. It extends from $30^{\circ}$ to $38^{\circ}$N latitude and from $-48^{\circ}$ to $-68^{\circ}$W longitude. The nomenclature used is north latitude is positive and west longitude is negative.

# II. LITERATURE REVIEW

Oceanography[1] is the study of the sea, embracing and integrating all knowledge pertaining to the sea's physical boundaries, the chemistry and physics of sea water, and marine biology. It encompasses portions of all the physical sciences. However, oceanography may be divided into five basic sciences: physics, chemistry, meteorology, biology, and geology. Of these five basic divisions, physical oceanography is the largest and most diversified. The physical aspect which will be investigated in this study is the determination of the speed of sound in sea water.

Inasmuch as characteristics of the sea may change with respect to both space and time, the periodicity and extent of these changes must be reviewed. Surveys are made at sea aboard a laboratory-equipped ship in order to obtain oceanographic information. The greater portion of oceanographic work at sea is carried out while occupying an oceanographic station in which the ship is lying-to (anchored). An oceanographic station[2] is any group of oceanographic observations made at the same, or virtually the same, geographic position at nearly the same time.

It is not a difficult matter to sample the surface of the ocean for chemical and physical analysis. But, it is more difficult to obtain a sample of water from the ocean deeps with accurate measures of depth and temperature and with the assurance that the sample is not mixed with surrounding waters.

Chemical properties of sea water are determined by analyzing samples of water obtained at various places and depths. Samples from below the surface are obtained by means of metal bottles designed for this purpose. The open bottles are attached at suitable intervals to a wire lowered into the sea. Each opened bottle flushes itself during lowering. When they reach the desired depths, a metal ring called a messenger is dropped down the wire. When the messenger arrives at the first bottle, it disconnects the top of the bottle from the wire. The bottle then reverses, making a 180$^{\circ}$ arc with the wire, while trapping a sample of the water at that depth; and releases a second messenger which travels on down the wire. The process is repeated at each bottle until all are closed, then they are hauled up and each bottle detached as it comes within reach. This widely used device is referred to by Bowditch[3] as the Nansen bottle and is shown in Figure 1. It is equipped with a removable frame for attaching thermometers which are shown in Figure 2.

If the oceans all over the world were sampled with the above techniques and the most accurate methods of chemical analysis for determining the composition of the sea water at each location were used, it would be found that, although the total amounts of dissolved salts is variable, the relative proportions of the major

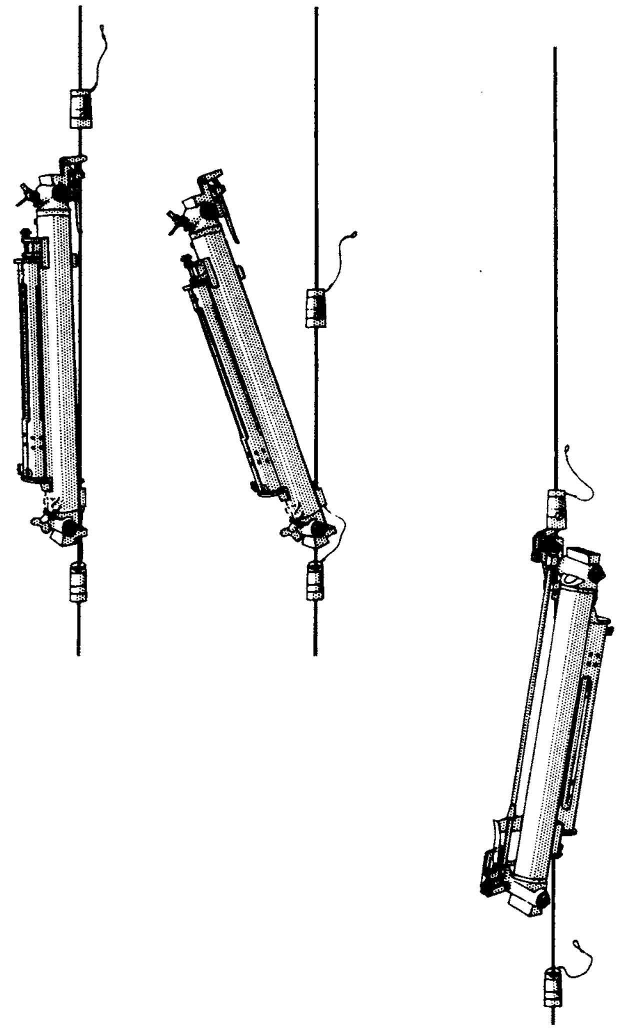



Figure 1. Nansen Bottle in Three Positions

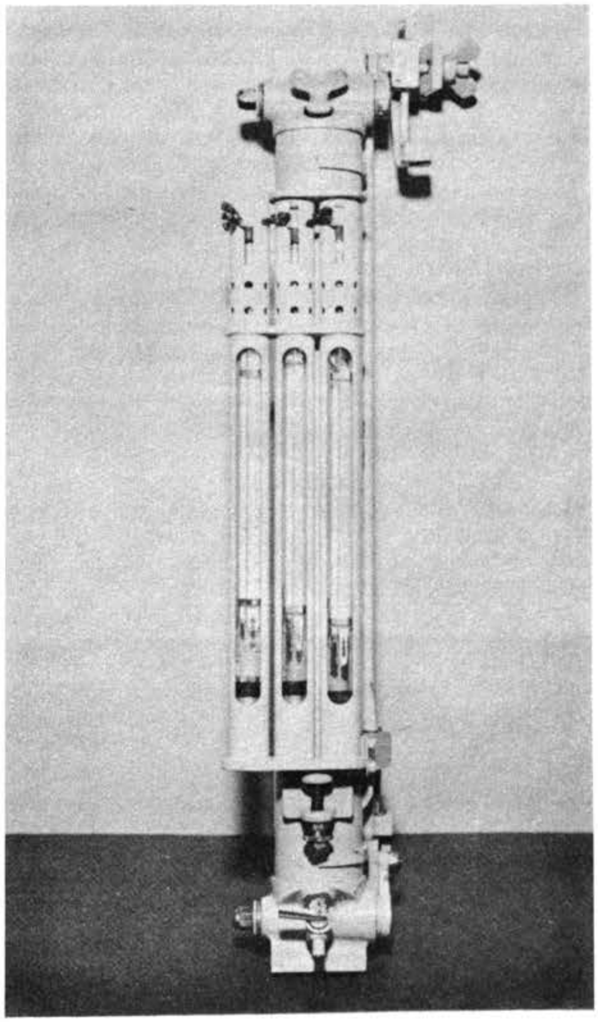


Figure 2. Nansen Bottle with Thermometers

elements (Na, Cl, Mg, Ca, etc.) are constant. Turekian[4] ascertained this fact from the study by W. Dittmar on the samples collected on the Challenger expedition in the 1870's.

The saltness of the ocean, or "salinity", is defined as the number of grams of dissolved salts in 1,000 grams of sea water. The total range of salinity of the open ocean is from 33 to 38 parts per thousand.

The oceans can be thought of as a gigantic pump which transfer heat from the equator to the poles. This transfer is effected in the surface waters of the oceans. The deep waters of the oceans all have their origins in the high latitudes and hence are considerably colder than the surface waters. The vertical distribution of temperature in the sea nearly everywhere shows a decrease of temperature with depth. Since colder water is denser, it sinks below warmer water. This results in a temperature distribution just opposite to that in the earth's crust, where temperature increases with depth below the surface of the ground. Figure 3 shows the variation of temperature and salinity with depth at one locality.

The peculiarities and limitations of sound in sea water could best be understood if the sound was transmitted in an ideal ocean. Such an ocean is assumed to be homogenous, infinite in extent, and perfectly elastic. It is obvious that the oceans have no such

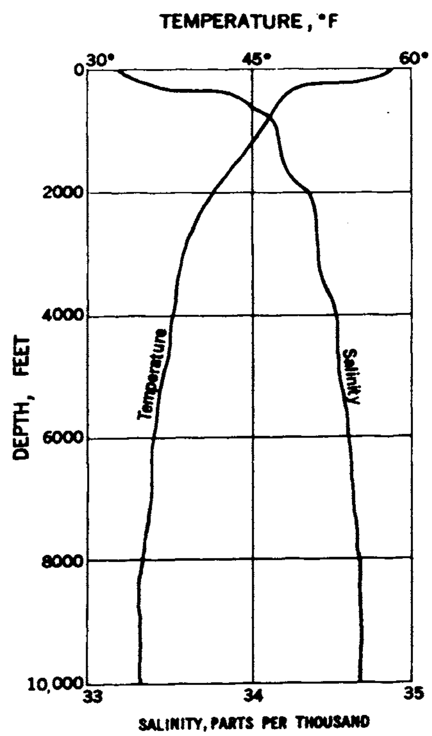



Figure 3. Variation of Temperature and Salinity with Depth at One Locality

ideal qualities. It is not infinite in extent, being bounded by the bottom and the surface. It is not homogenous; the upper layers are usually warmer than the lower ones. Furthermore, the properties of sound transmission are evidence of its non-elasticity.

Stephenson and Woodsmall[5] stated that the velocity of sound in water is given by the expression

$$V = \sqrt{dp/d\rho} \tag{2. 1}$$

where p is the pressure and $\rho$ is the density. Thus, the velocity of sound at a given point is determined by the pressure and density at that point. In the ocean, pressure increases with depth, and the density will vary as the temperature and salinity change. For convenience in calculation from well established physical data, the formula used is

$$V = \sqrt{r/\rho C} \tag{2. 2}$$

where r is the ratio of specific heats at constant volume and constant pressure, and C is the compressibility. The numerical values of these quantities depend on the temperature, salinity, and depth or pressure. Therefore, Berryman[6] concluded, the main environmental influences on sound velocity are temperature, salinity, and pressure.

In sea water an increase in temperature, pressure, or salinity results in greater speed of sound. Of the three, temperature has the greatest influence on the speed of sound in sea water in the upper layers. At depth, pressure, and in coastal areas, changes in salinity, may have the greatest effect. Normally, the change of these three elements is much more rapid in a vertical direction than in a horizontal direction. The change with depth varies with location. Typical curves showing the change of temperature and salinity with depth were shown in Figure 3. The increase of pressure with depth is almost uniform. A typical curve of speed of sound with depth is shown in Figure 4.

Sound energy moves through the water at speeds varying from less than 4,700 feet/second to greater than 5,100 feet/second. The first comprehensive sound velocity tables were published in 1924 by Heck and Service[7]. In 1939 Kuwahara[8] brought out a new set of sound velocity tables with added refinements and corrections for oceanographic work. Kuwahara presented sound speed as a function of temperature, pressure, and salinity. The tables of Kuwahara are believed to be the most accurate tables available at this time.

In general, sound velocity increases: (1) four to eight feet/second for every $1^{\circ}$F increase in temperature, (2) about four feet/second for every one part per thousand ($^{\circ}/_{oo}$) increase in salinity, and (3) about two feet/second for every 100 foot increase in

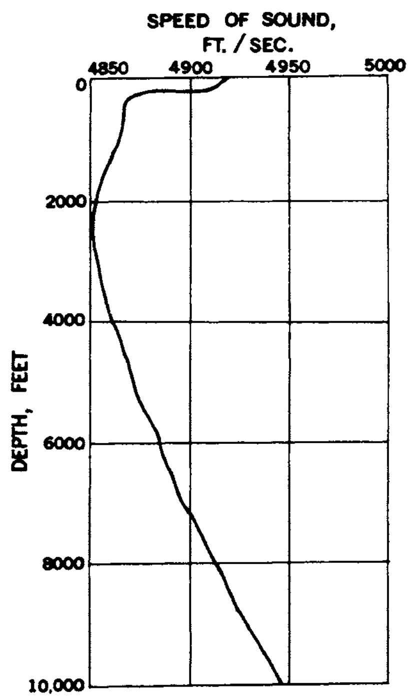



Figure 4. Typical Variation of Speed of Sound with Depth in the Ocean

depth[9]. Changes in the variables temperature, salinity, and depth or pressure are shown in Figure 5.

Kinsler and Frey[10] concluded that the composite effect of the factors temperature, salinity, and depth can be represented by the empirical equation

$$c = 1449 + 4.6t - 0.055t^2 + 0.0003t^3$$
$$+ (1.39 - 0.012t)(S - 35) + 0.017d \qquad (2.3)$$

where c is the velocity in meters per second, t is the temperature of the water in °C, S is its salinity expressed in parts per thousand, and d is the depth below the surface in meters. This equation is a simplification of more complicated equations obtained by Wilson[11] and others as the best fits to experimentally measured data.

Wilson's equation is given for the speed of sound in sea water. The experimental data used to determine this equation were obtained in temperature, pressure, and salinity ranges $-4°C < T < 30°C$, $1 kg/cm^2 < P < 1000 kg/cm^2$, and $0‰ < S < 37‰$. The equation is:

$$V = 1449.14 + V_T + V_P + V_S + V_{STP} \qquad (2.4)$$

$$V_T = 4.5721T - 4.4532X10^{-2}T^2 - 2.6045X10^{-4}T^3$$
$$+ 7.9851X10^{-6}T^4 \qquad (2.5)$$

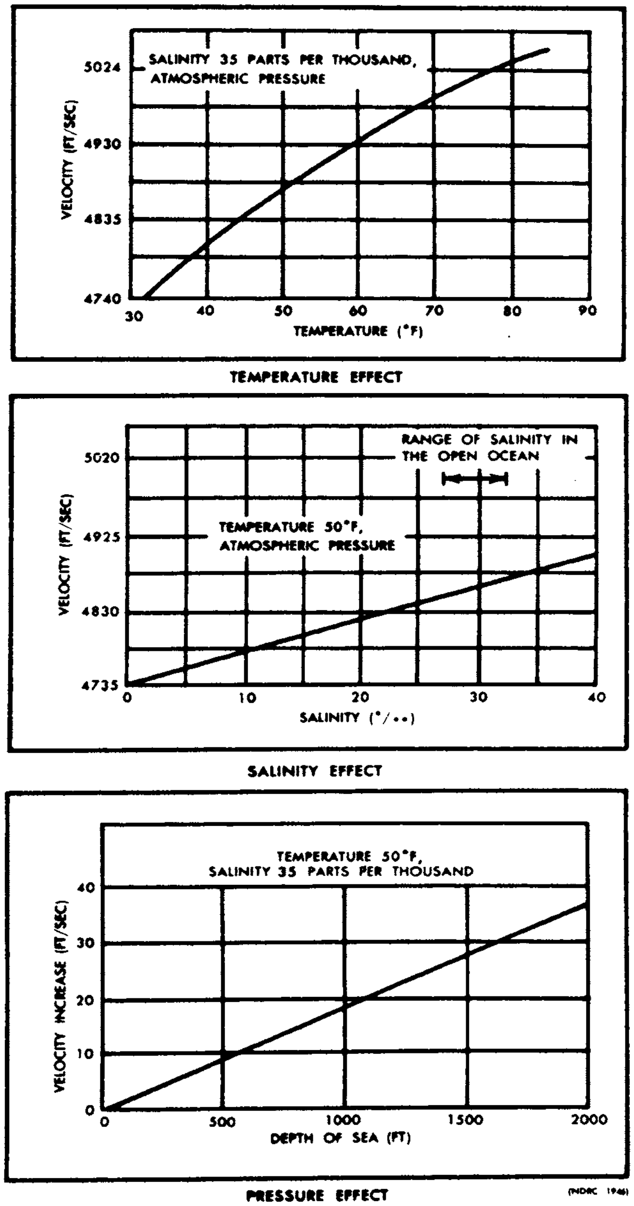



Figure 5. Effects of Variables on Sound Velocity

$$V_P = 1.60272\times10^{-1}P + 1.0268\times10^{-5}P^2 + 3.5216\times10^{-9}P^3$$

$$- 3.3603\times10^{-12}P^4 \qquad (2.6)$$

$$V_S = 1.39799(S - 35) + 1.69202\times10^{-3}(S - 35)^2 \qquad (2.7)$$

$$V_{STP} = (S - 35)(- 1.1244\times10^{-2}T + 7.7711\times10^{-7}T^2$$

$$+ 7.7016\times10^{-5}P - 1.2943\times10^{-7}P^2 + 3.1580\times10^{-8}PT$$

$$+ 1.5790\times10^{-9}PT^2 + P(- 1.8607\times10^{-4}T$$

$$+ 7.4812\times10^{-6}T^2 + 4.5283\times10^{-8}T^3)$$

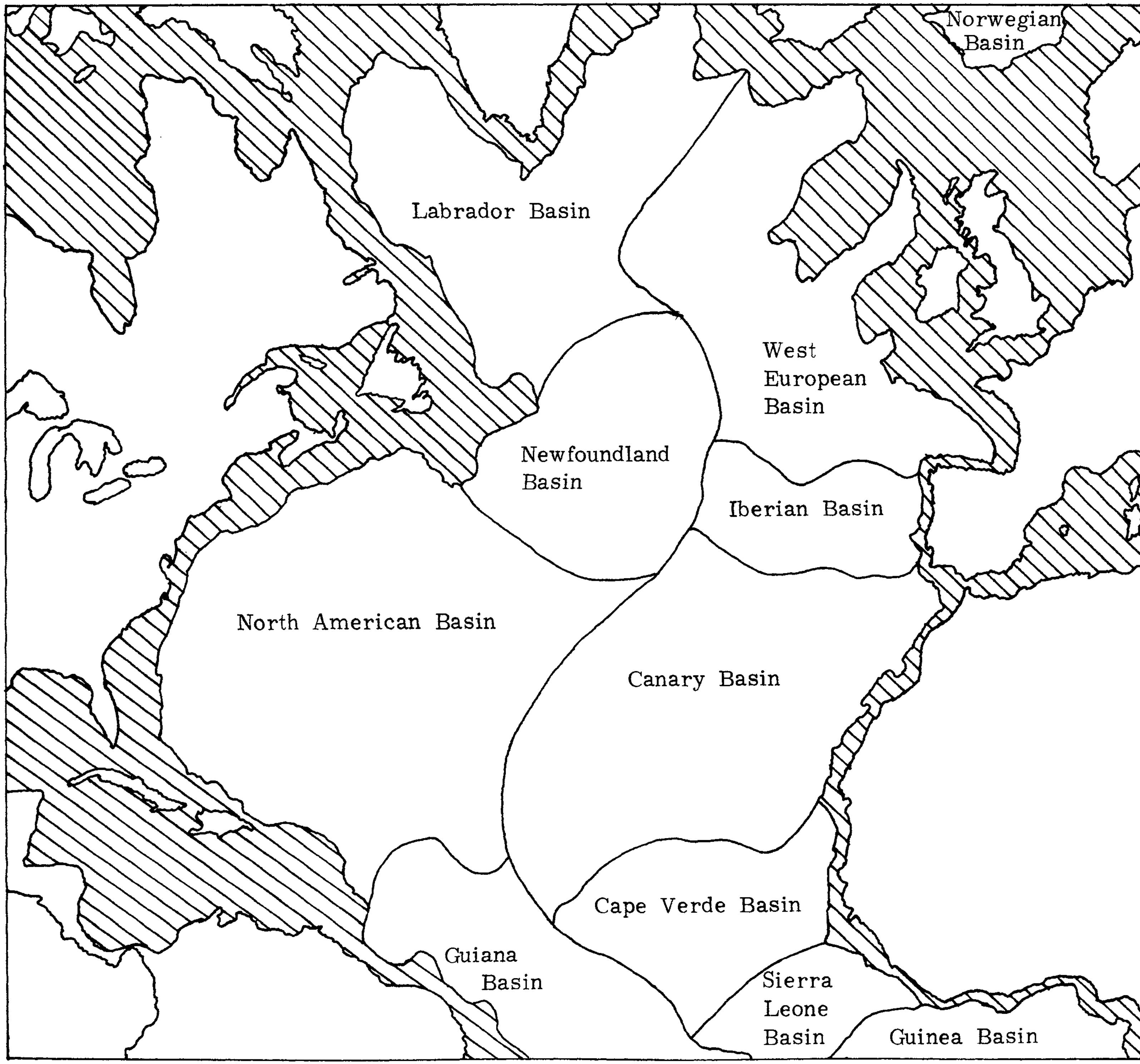
$$+ P^2(- 2.5294\times10^{-7}T + 1.8563\times10^{-9}T^2)$$

$$+ P^3(- 1.9646\times10^{-10}T). \qquad (2.8)$$

The units of temperature, pressure, salinity, and sound speed are °C, kg/cm$^2$, parts per thousand, and meters/second respectively. This equation may be used to predict sound speeds to within 0.30 m/sec over the indicated temperature, pressure, and salinity ranges.

Experimental multiple regression models have been investigated for a 4° by 4° square off the coast of New York in a study by Gillett[12]. In that study models were developed to predict ocean temperature, salinity, and sound velocity. The sound velocity model yielded a prediction equation which produced results that were nearly identical to values of Wilson's equation for that same area; plus the regression model was significantly simpler in form. It was established that sound velocity valid for only a very small area in the North Atlantic could be predicted adequately by regression equations. In his paper, he suggested that reasonable prediction equations could be obtained in larger areas of the North Atlantic Ocean. One of the purposes of this study is to pursue this idea and expand on its utility.

Generally, the speed of sound in the North Atlantic Ocean varies overall by less than five percent, its mean value being approximately 5,000 feet/second. As stated above, ocean sound speed has been demonstrated as a function of temperature, salinity, and pressure. Pressure produces a vertical gradient of approximately 0.016 feet/second/foot. On the other hand, temperature and salinity are the primary unpredictable variables which affect underwater sound speed. The natural variations in the temperature-salinity relationships of the major water masses occupying the deepwater physiographic basins below 1,000 fathoms (1 fathom = 6 feet = 1.83 meters) in the North Atlantic were investigated by Moore[13] to determine whether these variations resulted in appreciable differences in sound speed. The generalized outline of these basins is indicated by Map I; the shaded regions being the extensions of the continental landmasses which would appear




Map I. Generalize Outline of North Atlantic Ocean Basins if the present day sea level were lowered approximately 1,000 fathoms.

The major water masses in the North Atlantic below 1,000 fathoms are: North Atlantic Deep Water, North Atlantic Bottom Water, and South Atlantic Deep Water. The most pervasive water mass in the North Atlantic is the North Atlantic Deep Water, as it is located in all deep basins. Its temperature-salinity relation is almost linear, ranging roughly from approximately 3°C (37.4°F) at 1,000 fathoms to about 2°C (35.6°F) at its lower extent.

Differences between the temperature and salinity structures of the North Atlantic Deep and South Atlantic Deep Waters do not result in an appreciable change in average sound velocity. However, both water masses increase very slightly in temperature and salinity with increasing latitude, resulting in a very slight increase of sound speed at depth from south to north. In addition, sound speed is lower in the cooler, less saline North Atlantic Bottom Water than in the warmer, more saline North Atlantic Deep Water.

From the above discussion, it is apparent that sound velocity is more stable at the deeper underwater depths. It is this fact that led to this study. But, before a study such as this can be undertaken, a reliable source of data must be available. Through the cooperation of the United States Naval Oceanographic Office (NAVOCEANO), an atlas of the North Atlantic Ocean[14] was supplied along with numerous sound velocity articles. Originally over 100,000 oceanographic stations were available for that atlas. However, NAVOCEANO reduced the number of stations to a more tractable figure by eliminating undesirable or questionable data. Three criteria were established: (1) the oceanographic cast was required to extend at least to 100 meters' depth, (2) the surface measurements had to include both temperature and salinity, and (3) the deepest sound velocity value had to be greater than the value at the next-to-deepest depth in order to reject stations that did not extend to the depth at which minimum sound velocity occurs.

After applying the above three criteria, the output consisted of 22,832 observations. By an observation, in the atlas being considered, is meant the information necessary to draw and locate one sound velocity profile. Sound velocity information was computed from temperature, salinity, and depth data according to the equation developed by Wilson.

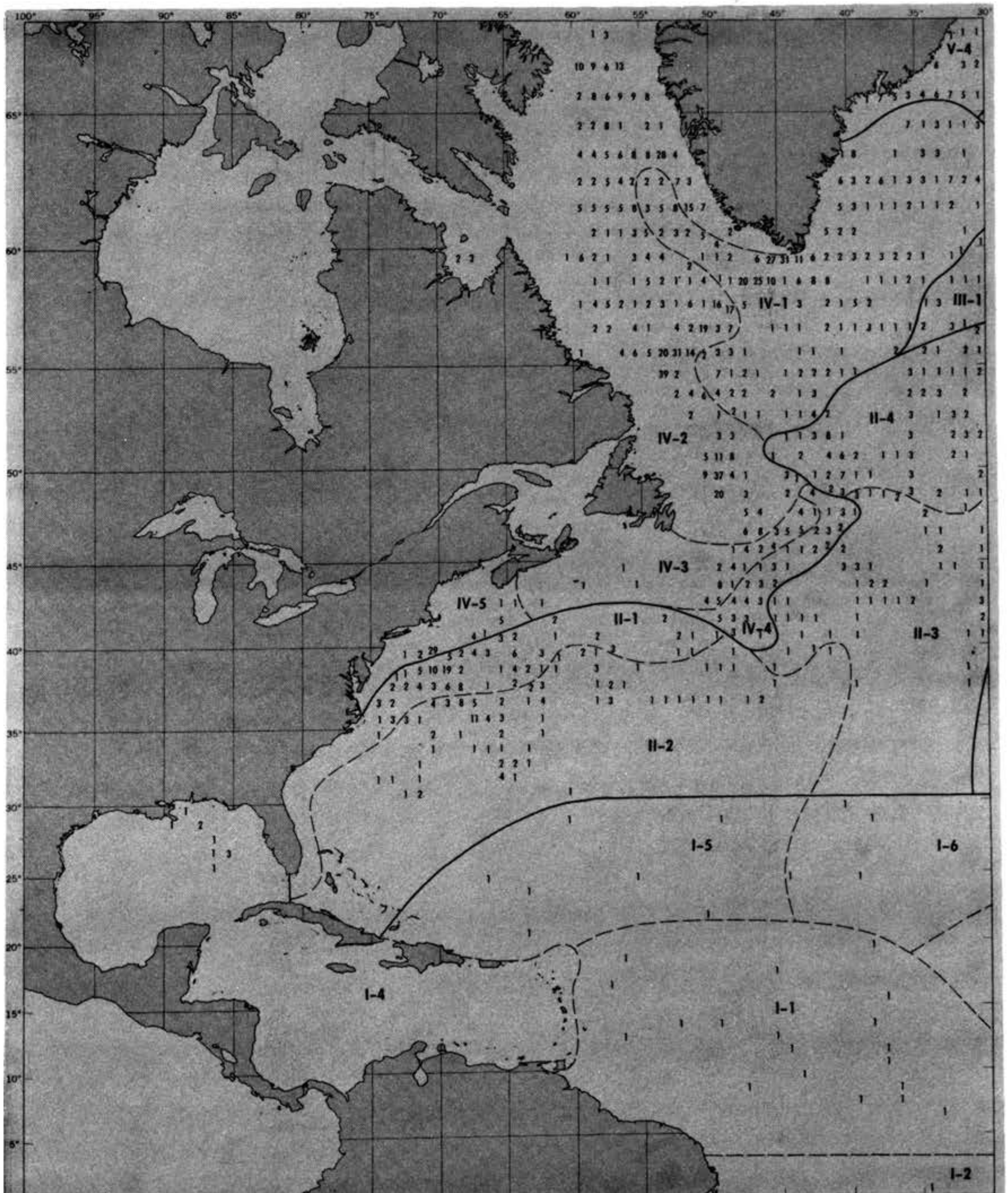
All observations were divided into seasons. The seasons were considered to consist of three month periods as follows: winter (January through March), spring (April through June), summer (July through September), and autumn (October through

December). The number of observations per season per one degree square was then plotted on four maps, one for each season. The purpose of the maps was to show the distribution of data in the North Atlantic Ocean. Map II illustrates the distribution of data for the summer season. The extent of the main geographic divisions are indicated by solid lines; subdivisions are outlined with dashed lines. The particular area of interest for this study lies in region II-2 of the map and is known as the Sargasso Sea of the Western North Atlantic.

From the sound velocity profiles, seasonal sound velocity profile envelopes were graphed for each region outlined on the data distribution maps. The winter, spring, summer, and autumn sound velocity profile envelopes for region II-2 are shown in Figures 6, 7, 8, and 9 respectively. In Figure 8 the shaded area indicates a subenvelope, representing the subrange of sound velocities within which a majority of the profiles lie. These four figures substantiate the idea of building more reasonable and feasible sound velocity models below some critical depth for large portions of the North Atlantic.

Table I gives the range of depths and sound velocities at the sound channel axis for the Sargasso Sea. Table II gives the number of profiles for the same area.

Station data of the atlas was originally stored on 75 tapes




Map II. Distribution of Data and Sound Velocity Structure Domains of the North Atlantic, Summer

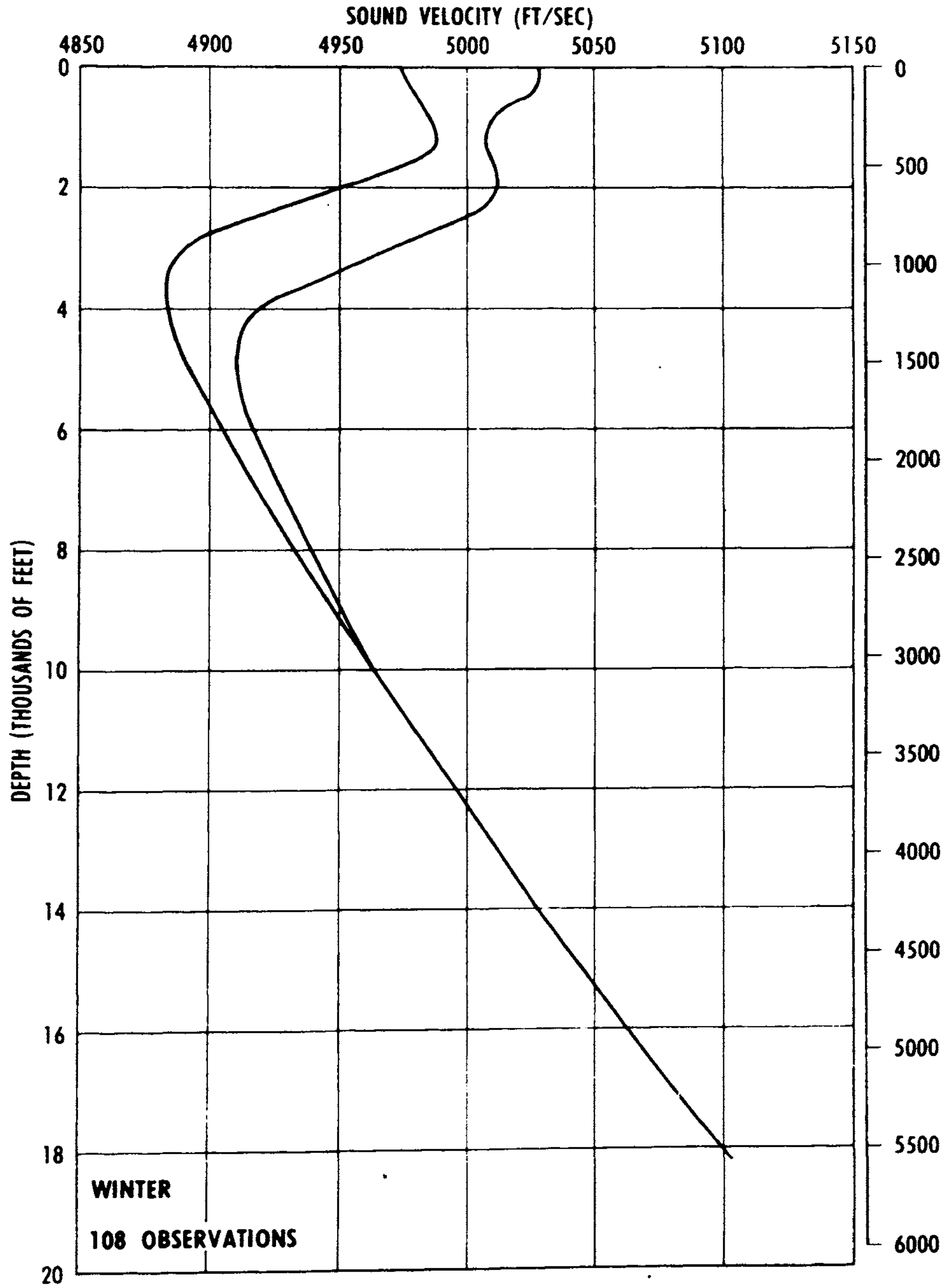



Figure 6. Sound Velocity Structure for Region II-2, Winter

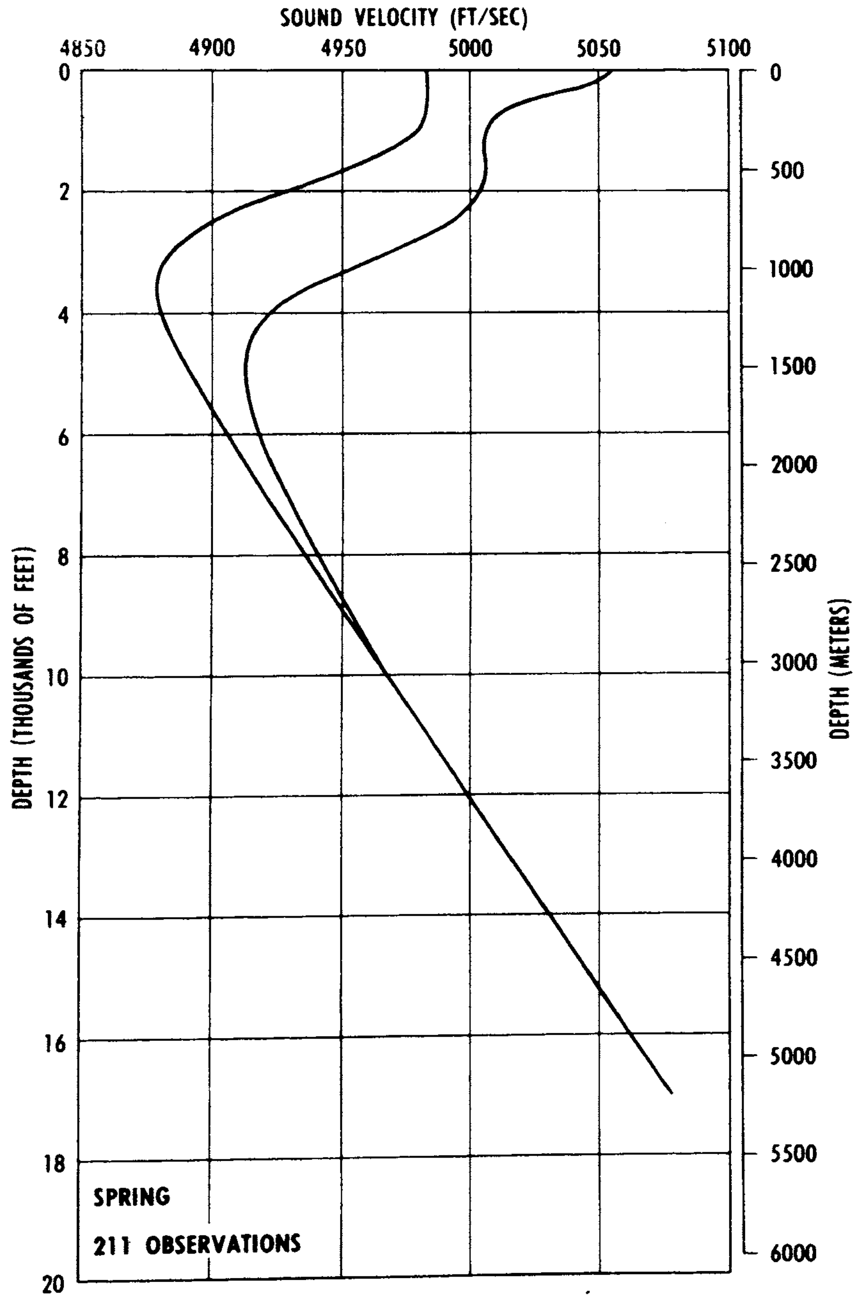



Figure 7. Sound Velocity Structure for Region II-2, Spring

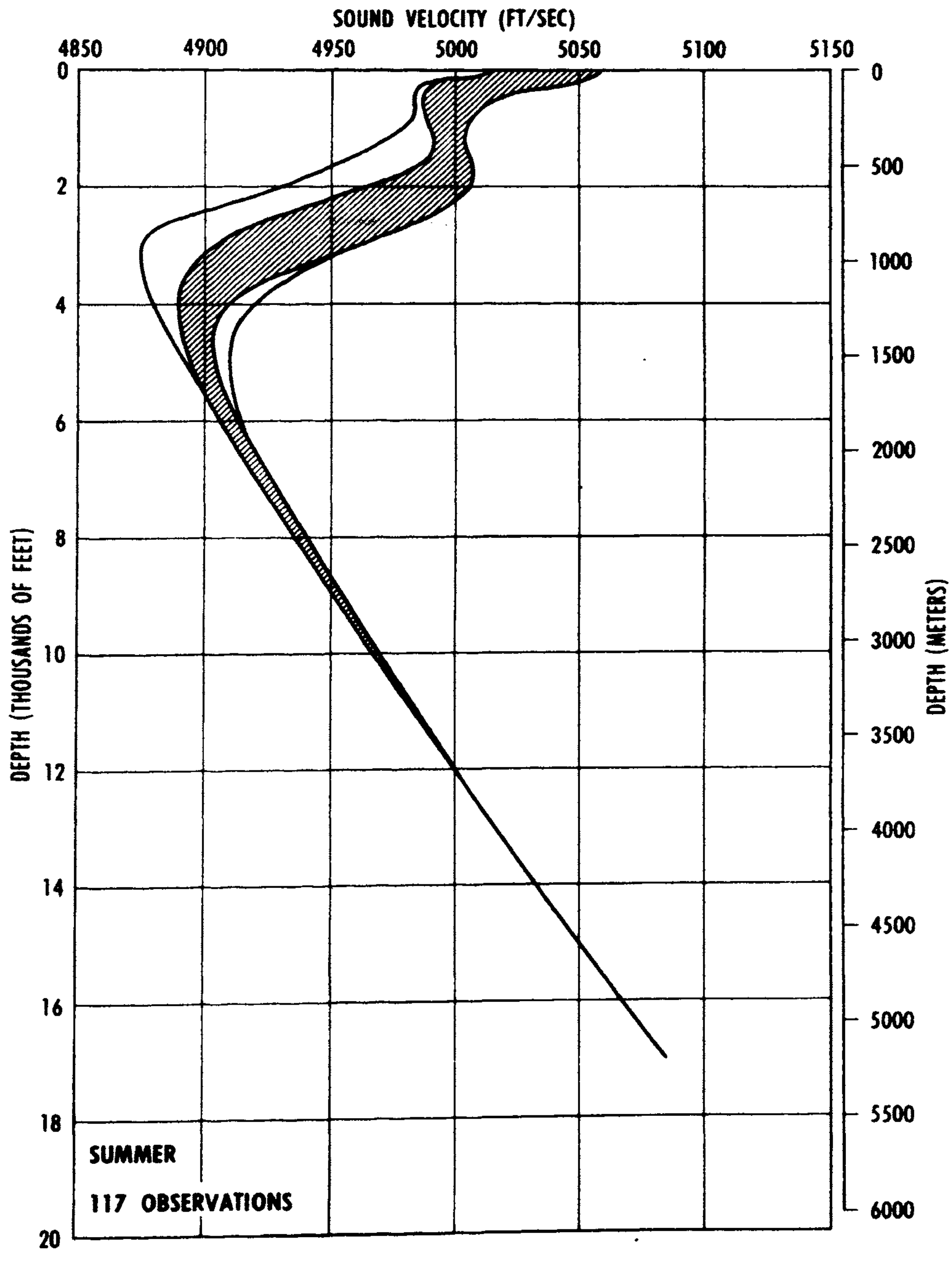



Figure 8. Sound Velocity Structure for Region II-2, Summer

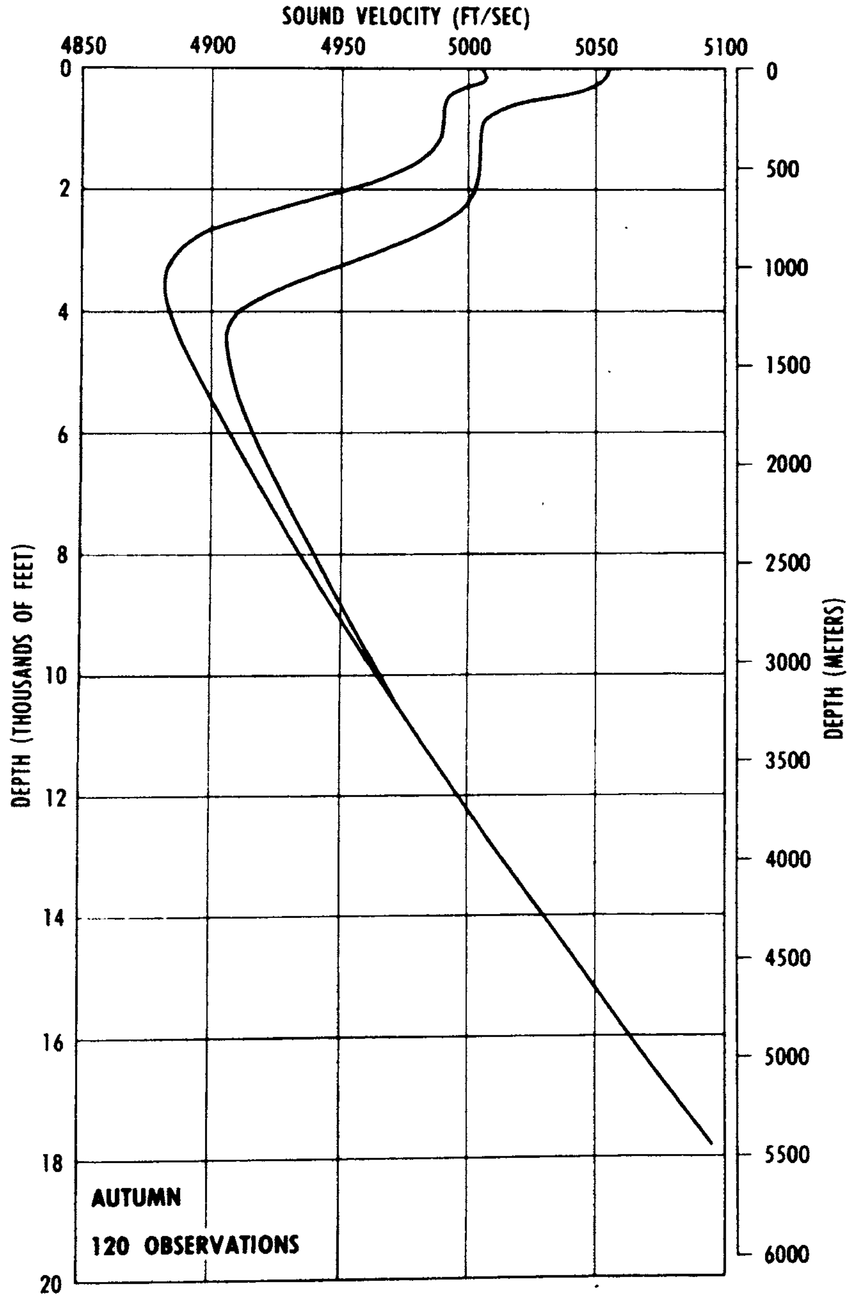



Figure 9. Sound Velocity Structure for Region II-2, Autumn

| SEASON | DEPTH (FT) MIN | DEPTH (FT) MAX | VELOCITY (FT/SEC) MIN | VELOCITY (FT/SEC) MAX |
|---|---|---|---|---|
| WINTER | 3600 | 4900 | 4885 | 4910 |
| SPRING | 3750 | 4900 | 4879 | 4913 |
| SUMMER | 3100 | 5200 | 4875 | 4910 |
| AUTUMN | 3600 | 4450 | 4881 | 4906 |

Table I. Range of Depths and Sound Velocities at the Sound Channel Axis in the Sargasso Sea of the Western North Atlantic

| SEASON | NUMBER OF PROFILES |
|---|---|
| WINTER | 108 |
| SPRING | 211 |
| SUMMER | 117 |
| AUTUMN | 120 |

Table II. Number of Profiles in the Sargasso Sea of the Western North Atlantic of card-line images. However, through the efforts of Yergen[15], the ocean station data is now stored on four "rapid access" tapes. The tape containing the data for the Sargasso Sea and surrounding area plus the program to access the tape were made available by NAVOCEANO.

# III. DISCUSSION

At present the highest accuracy achievable by computing sound speed from oceanographic station data consisting of observations of temperature and salinity at various depths is at best within a few feet per second. This is because of errors of measurement and systematic errors of the sound speed equation. Although sound velocity in the ocean is a function of depth, temperature, and salinity, there exists a pattern at the deeper depths in certain areas of the North Atlantic Ocean where sound velocity seems predictable as a function of other characteristics. Since the Sargasso Sea is so calm and of such a stable nature, it appears reasonable that a near homogenous population might exist with a fixed pattern. It is the objective of this study to find characteristics in the Sargasso Sea that would eliminate the need for on-location measurement of temperature and salinity to predict sound velocity without losing the accuracy of Wilson's equations. The characteristics investigated were seasonal factor, latitude, longitude, and depth.

In general no relationship exists between season, latitude, longitude, temperature, and salinity. However, some areas of the ocean, just because of the nature of the body of water, have established patterns of salinity and temperature with respect to position. Now the question arises whether the Sargasso Sea is an area where temperature and salinity have a predictable pattern or arrangement. With the assistance of graphic software it was displayed through a series of contour maps that temperature and salinity profiles at the deeper depths of the Sargasso Sea do have a predictable pattern that would allow sound velocity to be represented by a simple function. It was demonstrated that both temperature and salinity at a given depth plane change gently with respect to position to yield a smooth response surface. In addition, all the deeper depth planes have similar characteristics.

Adjacent to the west side of the Sargasso Sea is the Gulf Stream which rapidly shifts position depending on the season of the year. Consequently, the boundaries of the Sargasso Sea are constantly being altered. In order that a near homogenous area could be considered, the boundaries of the Sargasso Sea were investigated for all seasons. This investigation defined the area of interest as $30^{\circ}$ to $38^{\circ}$N latitude and $-48^{\circ}$ to $-68^{\circ}$W longitude. A diagram revealing the data distribution for this area is shown in Figure 10 at the depth plane of 2500 meters.

The utility of multiple regression analysis was explored in summarizing the entire body of collected data. The manner in which the observations were collected limits the amount of available data at the deeper depths. Only the best data available were used.

LATITUDE

38 37 36 35 34 33 32 31 30

| | | | | | | | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 1 | 1 | 2 | 2 | 2 | 3 | 0 | 0 | 0 | 1 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 0 |
| 2 | 0 | 3 | 0 | 1 | 2 | 0 | 0 | 0 | 1 | 2 | 0 | 0 | 0 | 1 | 1 | 1 | 1 | 1 | 0 |
| 8 | 11 | 2 | 0 | 1 | 2 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| 3 | 0 | 9 | 5 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 1 |
| 4 | 3 | 4 | 0 | 1 | 0 | 0 | 0 | 0 | 1 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| 0 | 2 | 3 | 2 | 2 | 1 | 1 | 1 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 2 | 0 | 1 |
| 0 | 2 | 3 | 3 | 1 | 0 | 0 | 0 | 0 | 0 | 1 | 1 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 1 |
| 2 | 0 | 3 | 1 | 1 | 1 | 0 | 1 | 1 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0 |

-68 -66 -64 -62 -60 -58 -56 -54 -52 -50 -48

LONGITUDE

Figure 10. Data Distribution at 2500 Meters

Appendix A contains the completely documented computer program activated on the IBM 360/50 computer and used for all regression analysis within this paper. In the program is developed statistical and computer techniques and methods for summarizing data to support Navy requirements in research. If n-1 equals the number of independent variables, then n is the number of independent variables plus the dependent variable. Thus, there will be n-1 simultaneous linear algebraic equations to solve. These equations, known as the normal equations, have a solution that usually results in handling a square matrix of order 2n-1 for stepwise multiple regression analysis procedures. However, through the efforts of Efroymson[16], this matrix can be handled in storage as an (n x n) square matrix. Thus, this allows an approximate reduction by a factor of four of core storage for a given number of independent variables. And, for a given size of core storage a greater number of independent variables are available. It should be noted that errors were encountered in the article by Efroymson and are corrected in Appendix A.

Although there are n-1 independent input variables, the number of independent variables possible for regression could become more numerous when other independent variables such as cross products are considered. To differentiate between an input variable and a regression variable, the first shall be referred to as a "variable" and the latter as a "term". The new techniques of independent term generation in the computer program is a significant contribution to the regression problem.

In the use of regression analysis it is necessary to know, or to assume, (1) the major independent variables, or main effects; and (2) a functional relationship between these variables, or a regression model. The oceanographic problem becomes one of searching for adequate models. It is indicated by graphic displays that adequate models can be derived for many ocean areas from the present archive of oceanic data. When working with regression analysis, the error between the predicted value and the actual value of the dependent variable for a particular observation is known as the residual ($e_i$ = residual for observation i). If m is the number of observations, then a regression equation is sought which will minimize $\sum_{i=1}^{m} e_i^2$. It is assumed that $e_i$ is a random variable with mean zero and some unknown variance $\sigma^2$, that is $E(e_i) = 0$, $V(e_i) = \sigma^2$. Furthermore, the assumption is made that $e_i$ and $e_j$ are uncorrelated, $i \neq j$, so that $cov(e_i, e_j) = 0$.

Let $c_i$ denote the true value of sound velocity in sea water for observation i, and let $\hat{c}_i$ denote the predicted value. Then the residual of observation i is

$$e_i = c_i - \hat{c}_i. \tag{3.1}$$

If this equation is juggled properly, according to Draper and Smith[17], it can be rewritten as

$$\sum_{i=1}^{m}(c_i - \bar{c})^2 = \sum_{i=1}^{m}(c_i - \hat{c}_i)^2 + \sum_{i=1}^{m}(\hat{c}_i - \bar{c})^2 \tag{3.2}$$

where $\bar{c}$ is the average sound velocity of the m observations. $c_i - \bar{c}$ is the deviation of observation i from the overall mean; and so, the left-hand side of equation (3.2) is the sum of squares of deviations of the observations from the mean. Since $c_i - \hat{c}_i$ is the deviation of the ith observation from its predicted or fitted value, and $\hat{c}_i - \bar{c}$ is the deviation of the predicted value of the ith observation from the mean, equation (3.2) can be expressed in the following form.

$$\begin{matrix}\text{Sum of squares} \\ \text{about} \\ \text{the mean}\end{matrix} = \begin{matrix}\text{Sum of squares} \\ \text{about} \\ \text{regression}\end{matrix} + \begin{matrix}\text{Sum of squares} \\ \text{due} \\ \text{regression}\end{matrix} \tag{3.3}$$

One measure of the goodness of fit of a regression equation is the square of the multiple correlation coefficient ($R^2$). By definition the ratio

$$R^2 = \frac{\text{Sum of squares due to regression}}{\text{Sum of squares about the mean}} = \frac{\sum(\hat{c}_i - \bar{c})^2}{\sum(c_i - \bar{c})^2}. \tag{3.4}$$

$R^2$ is a measure of the usefulness of the regression terms, other than the constant term, in the model. If the prediction is perfect, then $R^2 = 1$. $R^2(100)$ is the percent of the total variation about the mean ($\bar{c}$) that is accounted for by regression. Ideally a suitable regression model is sought that has the least number of terms as possible with a high $R^2$ as close to one as possible.

All ocean station data used within this study were collected at specific depth planes. Each observation, beginning at 500 meters and extending to the ocean floor, had a recorded depth of 500, 600, 800, 1000, 1200, 1500, 2000, 2500, 3000, 4000, or 5000 meters. The data contained on the magnetic computer tape had a major sort on the depth planes mentioned above. Each observation contained the following information: (1) day of month, (2) month, (3) year, (4) hour, (5) latitude in thousandths of a degree, (6) longitude in thousandths of a degree, (7) depth in meters, (8) temperature in degrees centigrade, (9) salinity in parts per thousand, (10) $\sigma_t$, (11) sound velocity in feet per second, (12) oxygen content, (13) quality code, and (14) bottom depth. Table III exhibits the average temperature and salinity plus the number of observations and average sound velocity at each depth plane from 500 meters to the ocean floor.

The accuracy obtained by Wilson's sound velocity equations is also the accuracy required in this study if its objectives are to be fulfilled. In other words, a successful model must have a standard deviation of sound velocity of approximately 1 ft/sec or

| DEPTH PLANE | AVERAGE TEMPERATURE | AVERAGE SALINITY | AVERAGE SOUND VELOCITY | NUMBER OF OBSERVATIONS |
|---|---|---|---|---|
| 500 | 16.6838 | 36.2895 | 4991.312 | 336 |
| 600 | 15.3909 | 36.0725 | 4985.699 | 333 |
| 800 | 11.3923 | 35.4947 | 4950.652 | 328 |
| 1000 | 7.5268 | 35.1401 | 4913.566 | 316 |
| 1200 | 5.4328 | 35.0594 | 4897.093 | 308 |
| 1500 | 4.3490 | 35.0109 | 4898.855 | 261 |
| 2000 | 3.6726 | 34.9779 | 4917.214 | 182 |
| 2500 | 3.3108 | 34.9686 | 4940.187 | 130 |
| 3000 | 2.9165 | 34.9507 | 4962.890 | 95 |
| 4000 | 2.3615 | 34.9079 | 5012.273 | 60 |
| 5000 | 2.2330 | 34.8860 | 5068.609 | 10 |

Table III. Ocean Station Data at Specified Depth Planes less with a reasonable number of independent terms that are not dependent on data sampling.

A review of Figures 6, 7, 8, and 9 in Chapter II suggested that 3000 meters should be the maximum critical depth considered. A simplified regression model of sound velocity within the specified accuracy was easily obtained for depth planes of 3000 meters and below. Essentially, the independent term of depth controlled the model. This expected result warranted an investigation of 2500 meters as the critical depth.

Since the sound velocity profiles begin to show some scattering at a depth of 2500 meters, it was puzzling whether season had any accountable significance in this behavior. Graphs of month vs temperature, month vs salinity, and month vs sound velocity showed no predictable pattern. This strongly indicated that season would not be a significant factor of sound velocity below 2500 meters in the Sargasso Sea. To reassure this conclusion, month was considered as one of the possible input variables with the hope that it would be rejected (not enter in regression).

In an effort to locate all independent terms which have the slightest correlation with sound velocity as a function of position, many regression models with a low F-level of 1.25 were investigated in a systematic manner with the sole purpose of collecting a library of terms. The input variables considered in building this library of independent terms were month, latitude, longitude, and depth and at times will be referred to as M, L, LN, and D respectively. In addition sound velocity will be denoted by SV.

The first part of this systematic approach involved building a regression model for each combination of the four input variables taking two at a time with sound velocity as the dependent variable. Thus, the number of computer runs executed was six (4!/2!2!). Each computer run had 99 independent terms available for regression with the highest power on any particular input variable not exceeding nine. Equivalently, this represents all possible combinations of the two input variables being considered. These computer runs were accomplished by use of NORMAL mode as explained in Appendix A. Next a regression model for each combination of the four input variables taking three at a time was implemented with sound velocity as the dependent variable and resulted in four (4!/3!1!) runs. Each computer run had 99 independent terms available for regression with three as the highest power on some input variables and four on others. Again these computer runs were carried out as defined by NORMAL mode. Finally, all four input variables were considered in a single regression model with sound velocity as the dependent variable. This regression model had 80 independent terms available for regression under NORMAL mode with the highest power of any given input variable being two.

After recalling the contour maps previously discussed, the decision was made that the behavior of all input variables would not justify any independent terms in the above computer runs with a power larger than five. When re-examining all the independent terms with powers higher than five, it was found that these terms entered at a low F-level with neither significantly decreasing the standard error of sound velocity nor increasing the value of $R^2$. In fact, many times no change was evident by the addition of one of these terms.

After collecting all independent terms that entered into regression from all above computer runs and eliminating the terms with powers higher than five, the following independent terms comprise the library of terms: D, $D^2$, $D^3$, $D^4$, $(LN)D^3$, $(LN)^4$, $(LN)^4D^4$, L, $LD^3$, $L(LN)^2$, $L^2(LN)D$, $L^4$, $L^4D$, $L^4D^5$, M, MD, M(LN), $M(LN)^4D^2$, $M^2D^5$, $M^2(LN)^4D^2$, $M^3$, $M^3D$, $M^3D^2$, $M^3D^3$, $M^3(LN)^4D^2$, $M^3L^2(LN)$, $M^3L^4(LN)^4$, $M^4L^3$, and $M^5D$. One should keep in mind that this library has many insignificant independent terms because of the low F-level specified in the regression program plus the massive term collecting procedures.

When permitting the entire library of terms to become available as independent terms for regression, many excellent models were established through the use of the ACCEPT mode in the computer program. However, keeping in mind the objectives of this study, the most satisfying model was

$$SV = 4803.75202 + 4.134663X10^{-3}D + 1.033088L$$

$$+ 8.623X10^{-12}LD^3 - 1.948X10^{-9}L^4D \qquad (3.5)$$

with the standard deviation of sound velocity from the supplied data being 1.02 ft/sec and the multiple correlation coefficient squared having a value of 0.9991 or 99.91%. The order of entry for each independent term plus its contribution to reducing the standard error of sound velocity and increasing $R^2$ are shown in Table IV.

As previously suspected, this model rejected a seasonal factor by not allowing any independent term that was a function of month to enter regression. However, if sound velocity is to be a function of position, it is somewhat surprising that none of the independent terms which were a function of longitude entered regression. It should be noted that the next term that would have entered into the regression model was a function of longitude. The validity of the independent terms in equation (3.5) will be discussed later in this chapter.

This model is satisfying when considering its simplicity, the accuracy inherent in the low standard error and high $R^2$, the large area of interest, and the elimination of on-location data

| STEP NUMBER | INDEPENDENT TERM | STANDARD ERROR OF SOUND VELOCITY | MULTIPLE CORRELATION COEFFICIENT SQUARED (%) |
|---|---|---|---|
| 1 | D | 1.9469 | 99.65% |
| 2 | $LD^3$ | 1.1189 | 99.89% |
| 3 | L | 1.0981 | 99.89% |
| 4 | $L^4D$ | 1.0200 | 99.91% |

Table IV. Variables Entering Regression

collection. Nevertheless, a different approach was tested in an attempt to improve the above model for the same critical depth of 2500 meters.

Since temperature and salinity are gently changing within the deeper depth planes, it was thought that an average temperature and salinity at the various depth planes, along with the variables latitude and depth, might yield a more precise model. The average values of temperature and salinity employed at the respective depth planes are listed in Table III and shall be referred to as $T_c$ and $S_c$ respectively.

Again the same systematic approach was utilized with respect to the input variables of constant temperature, constant salinity, and depth. In other words, the first three computer runs consisted of building a regression model for each combination of the three input variables taking two of them at a time. Then all three input variables were made available in a single computer run.

After reviewing all four computer runs, no new terms were found to be significant. In a final effort, the input variables latitude, depth, constant temperature, and constant salinity were combined into a single computer run. The following model resulted.

$$SV = 4705.30155 + 0.48452 T_c S_c + 5.887591 \times 10^{-2} D + 1.042915 L \tag{3.6}$$

This model has a standard deviation of sound velocity of 1.0256 ft/sec and a $R^2$ of 99.90%. Although this regression equation is extremely simple, it has the disadvantage of having two additional input variables ($T_c$ and $S_c$) whose values are a function of the existing data. Since this model did not improve either the standard deviation or $R^2$, it can be concluded that the model of equation (3.5) is more profitable.

It was reasonable to now conclude that the critical depth can be raised from 3000 meters to 2500 meters with the model represented by equation (3.5). However, the question should be asked whether it is feasible for 2000 meters to become the critical depth. An investigation of 2000 meters as the critical depth led to the following results: (1) standard deviation of sound velocity of approximately 1.25 ft/sec, and (2) multiple regression correlation coefficient of approximately 99.88%. Since this is not within a reasonable limit of the desired accuracy, it can now be concluded that the critical depth as established by the criteria of this study is 2500 meters; and, the best model below this critical depth is represented by equation (3.5).

The precision of the sound velocity model depends upon the precision of the supplied data and upon the fit of the model to the data. The difference between the computed data and supplied data may be used to estimate the precision of the results. To further explore the utility of equation (3.5), Figure 11 illustrates the distribution of the residuals for the 295 observations utilized in this study. Furthermore, 139 residuals had a value less than zero with the remaining 156 residuals being a value greater than zero. The longest run of residuals with the same sign was 38 at the depth plane of 3000 meters.

Regression analysis appears to have considerable potential as a technique for estimating sound velocity. However, the physical reality of the estimates must also be considered, since it is always possible to improve the statistical measures of goodness of a regression-model estimate by merely adding additional terms to the model. From a physical viewpoint these terms may be nonsense terms. Therefore, some justification of the terms in equation (3.5) will follow.

Deep water is produced by the cooling of the surface water at high latitudes in late winter. This cooling increases the density of the surface water to the point at which it is able to sink to the bottom and flow southward filling the deep basins. In the Atlantic Ocean there is a large water mass known as the North Atlantic Deep Water which comprises about half the total volume of the ocean and which is clearly of North Atlantic origin. This fact suggests the investigation of the change of temperature and density with latitude.

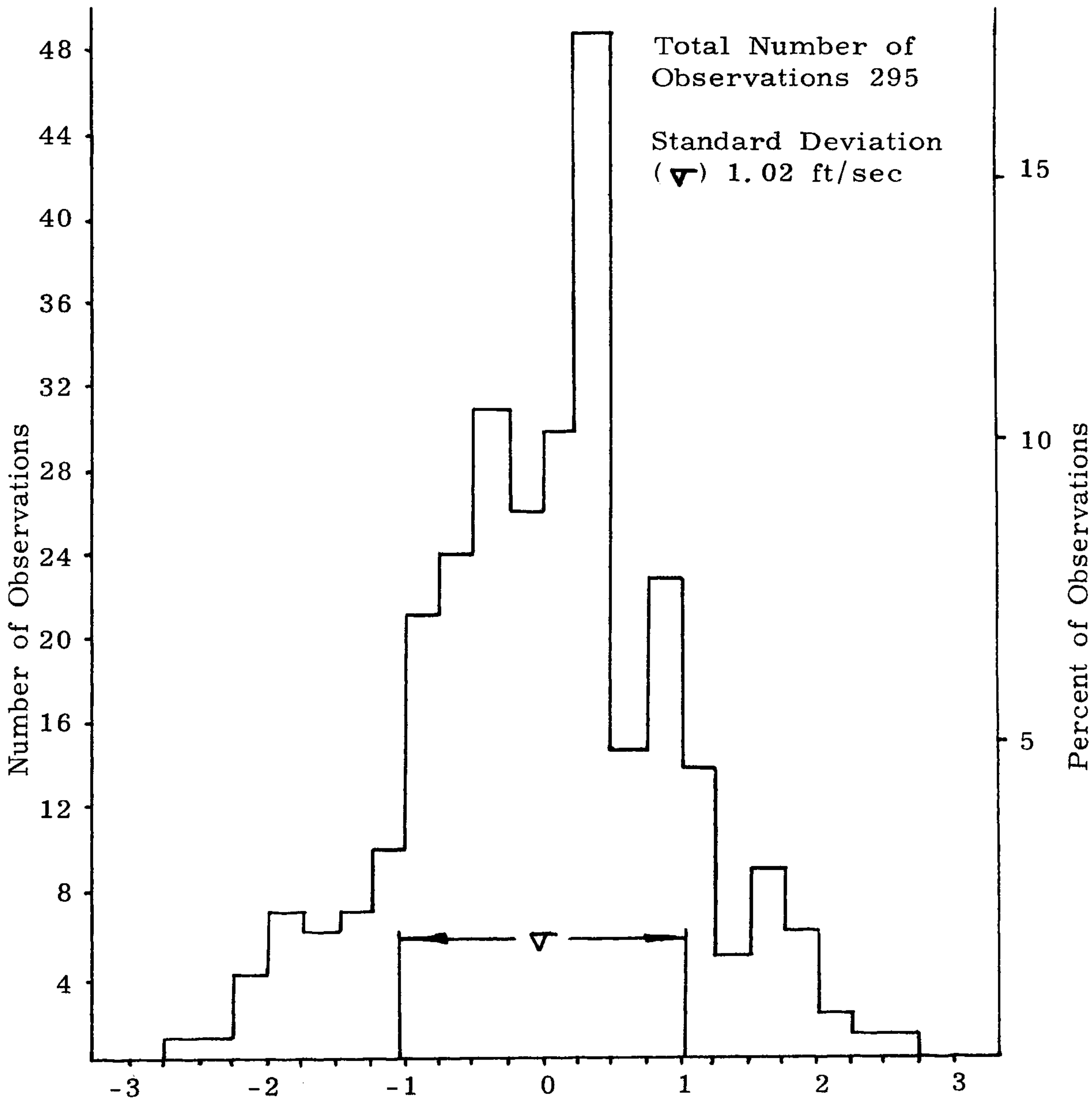



Figure 11. Sound Velocity Residuals, ft/sec

Many scattergram plots were examined for this particular area of the Sargasso Sea at the depth plane of 2500 meters. Two rewarding results were discovered: (1) a definite increase in sound velocity with an increase in temperature that appeared close to being linear, and (2) a slight increase in sound velocity with an increase in latitude. These results prompted the investigation of latitude with respect to temperature. The area of interest at a depth of 2500 meters was divided into strips, each $1^{\circ}$ of longitude in width and eight degrees of latitude in height. Another way to describe this division would be to let each column of Figure 10 become equivalent to one strip.

A scattergram plot of latitude vs temperature was made for each strip. A typical scattergram plot is illustrated in Figure 12 for a constant longitude of $-65^{\circ}$. After combining all the information obtained from the plots, it was found that an increase of latitude in a northern direction resulted in an increase in temperature which also implies an increase in sound velocity. But, this is not surprising, for, as Moore[13] pointed out, the North Atlantic Deep Water increases very slightly in temperature and salinity with increasing latitude, resulting in a very slight increase of sound velocity at depth from south to north. This fact supports predicting sound velocity as a function of latitude as is done in equation (3.5). A scattergram plot revealed that an increase in

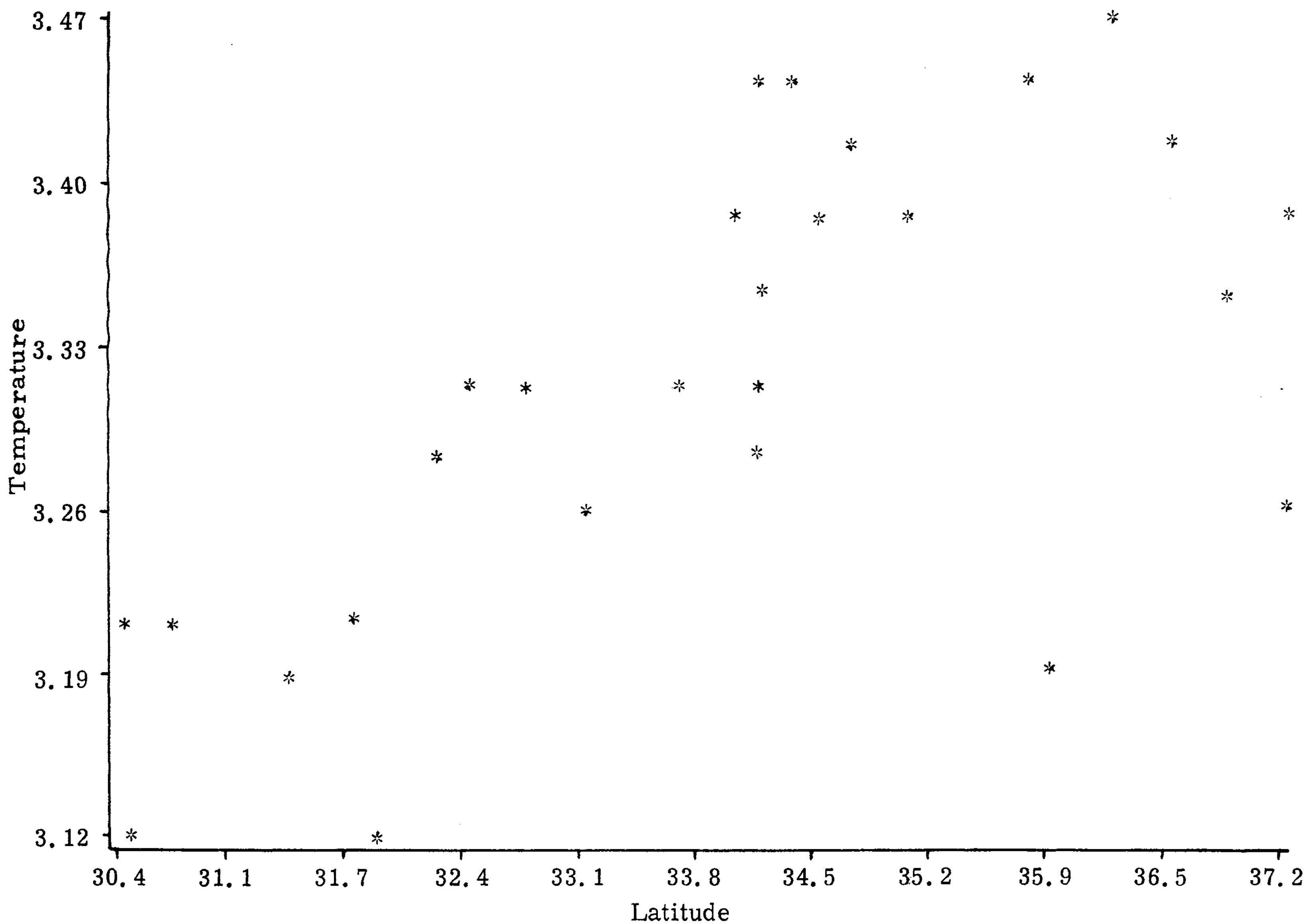



Figure 12. Scatter Diagram of Latitude vs Temperature

depth results in a close to linear increase in sound velocity.

Of course, depth as a linear independent term needs no justification since in equation (2.3) Kinsler and Frey[10] have established sound velocity as a linear function of depth. However, since salinity contributes such a small amount to this equation, it could be rewritten as

$$SV = k + aD + f(t) \tag{3.7}$$

where k is a constant and a is the coefficient of the linear independent term of depth. This leaves f(t) to be established.

Using temperature as the dependent variable, a regression model was developed with latitude and depth as input variables. This computer run demonstrated temperature for the area of interest to be a highly correlated function of the independent terms D, L, and $LD^3$. Thus, equation (3.7) nearly takes the form of equation (3.5). The only independent term not appearing is $L^4D$; but, it was the last or least correlated independent term that entered into regression equation (3.5). $L^4D$ might account for the part of equation (2.3) ignored by estimating it by equation (3.7).

For use in oceanographic work it is generally more convenient to express sound velocity in terms of depth instead of pressure. The conversion of pressure to an equivalent depth may be done by dividing the depth into thin layers and summing the product of the average density in each layer and the thickness of each layer. Wilson[18] formulated this numerical integration as $P_i = \sum \overline{g}_\Theta \overline{\rho}_i t$ where $P_i$ is the pressure at the bottom of layer i, $\overline{\rho}_i$ is the average density between each layer, t is the thickness of the layer, and $\overline{g}_\Theta$ is the acceleration due to gravity at the latitude $\Theta$ and the mean depth of the layer. Since sound velocity is a function of pressure, the above numerical integration further substantiates the validity of the sound velocity model represented by equation (3.5) as a function of depth and latitude. However, since latitude is present in the sound velocity model, and it is a function of both temperature and gravity, the relative amount of latitude accredited to temperature and gravity respectively is undetectable.

It was concluded that the regression model, as expressed in equation (3.5), is a physically acceptable estimator of sound velocity for the area covered by these analyses.

## IV. CONCLUSION

In this study the desired functional relationship of sound velocity in terms of position only was determined from analytical and theoretical considerations and from an investigation of scatter diagrams prepared from the data being analyzed. A regression model of sound velocity was developed consisting of latitude and depth that is capable of estimating sound velocity below a critical depth in a large area of the North Atlantic Ocean to a standard deviation of approximately 1 ft/sec. The critical depth established for the desired accuracy was found to be 2500 meters.

From both a statistical and a physical viewpoint, relatively simple regression models have a considerable potential as estimators of sound velocity with a high degree of accuracy at the deeper depth planes. It is assumed such models can be derived for many ocean areas utilizing the present archive of ocean station data.

On the assumption that the regression model is reasonably valid, the regression technique has the capability of eliminating the collection of ocean station data at the deeper depths and, thus, resulting in a substantial saving. In addition, it provides an effective method of editing sound velocity data for erroneous observations.

This study suggests regression techniques may be used as a new approach to summarizing archived ocean station data that is more objective and amenable to computer usage than present methods.

As an outgrowth of this study it is recommended that regression models be investigated in other areas of the ocean with some correlation between accuracy and critical depth. When shallower critical depths are considered, seasonal effects could become a significant factor.

## V. BIBLIOGRAPHY


1. Baker, B. B., Jr., W. R. Deebel, and R. D. Geisenderfer, editors. Glossary of Oceanographic Terms, second edition. Oceanographic Analysis Division, Marine Sciences Department, U. S. Naval Oceanographic Office, Special Publication 35, Washington, D. C., 1966.

2. Instruction Manual for Obtaining Oceanographic Data, third edition. U. S. Naval Oceanographic Office, Publication Number 607, Washington, D. C., 1968.

3. Bowditch. Oceanography, Part Six of American Practical Navigator (1958). U. S. Navy Hydrographic Office, Publication Number 9, 1966.

4. Turekian, Karl K. Oceans. Englewood Cliffs: Prentice Hall, Inc., 1968.

5. Stephenson, E. B. and F. J. Woodsmall. Report on the Velocity of Sound in Sea Water. Naval Research Laboratory, Anacosta Station, NRL Report Number S-1722, Washington, D. C., 1941.

6. Berryman, Matilene S. Lecture Notes on Basic Underwater Sound. Naval Oceanographic Office, Informal Report Number 67-54, Washington, D. C., 1967.

7. Heck, N. H. and J. H. Service. Velocity of Sound in Sea Water. Coast and Geodetic Survey, Special Publication Number 108, 1924.

8. Kuwahara, S. "Velocity of Sound in Sea Water and Calculation of the Velocity for Use in Sonic Sounding." Hydrographic Review, Volume XVI, Number 2, 1939.

9. Oceanography and Underwater Sound for Naval Applications. U. S. Naval Oceanographic Office, Special Publication 84, Washington, D. C., 1965.

10. Kinsler, Lawrence E. and Austin R. Frey. Fundamentals of Acoustics, second edition. John Wiley and Sons, Inc., 1967.

11. Wilson, Wayne D. "Equation for the Speed of Sound in Sea Water," Journal of the Acoustical Society of America, volume 32, number 10, page 1357, October 1960.

12. Gillett, Billy E. Critical Study of Approximating Functions (and Methods) as Applied to Ocean Station Data - Project I Regression Analysis. University of Missouri-Rolla, Rolla, Missouri, 1968.

13. Moore, Carl T. Tables of Sound Speed at the Bottom below 1,000 Fathoms in the North Atlantic Ocean. Marine Sciences Department, U.S. Naval Oceanographic Office, Washington, D.C., 1965.

14. Oceanographic Atlas of the North Atlantic Ocean, Section VI, Sound Velocity. U.S. Naval Oceanographic Office, Publication Number 700, Washington, D.C., 1967.

15. Yergen, Walter E. A Rapid Access Tape Format for Oceanographic Station Data, Research and Development Department, U.S. Naval Oceanographic Office, Informal Report Number 67-14, Washington, D.C., 1967.

16. Efroymson, M. A. "Multiple Regression Analysis." Mathematical Methods for Digital Computers, Volume 1, edited by Anthony Ralston and Herbert S. Wilf. Wiley, 1960.

17. Draper, N. D. and Smith, H. Applied Regression Analysis. John Wiley and Sons, Inc., 1968.

18. Wilson, Wayne D. Journal of the Acoustical Society of America, volume 32, number 6, pages 641-644, June 1960.

# VI. VITA

Richard Roland Kunkel was born June 5, 1945 in St. Louis, Missouri. His primary and secondary education was also received in St. Louis. His secondary education was highlighted by his attendance at Missouri Boy's State.

In September 1963 he started his undergraduate training upon entering the University of Missouri - Rolla. Before receiving his Bachelor of Science degree in Applied Mathematics in May 1967, he had experienced membership and offices in honorary and professional societies.

Initially enrolled in the Graduate School of the University of Missouri - Rolla in June 1967, he maintained the Seismic Station of the university in fulfillment of his assistantship for the summer of 1967. In September 1967 a National Science Foundation Grant was awarded to him. He received his Master of Science degree in Geophysical Engineering in January 1969. He continued to hold the National Science Foundation Grant while working toward his Doctor of Philosophy degree in Geophysical Engineering.

His professional experience includes working as an Industrial Engineer during the summer of 1966 for United States Steel Corporation and as an Oceanographer for the United States Naval Oceanographic Office during the summer of 1969. It should be noted that he has been associated with naval projects since September 1967.

On June 3, 1967 he was married to the former Karen Frances Short. At the time of this writing, they have one child, a boy named Timothy Paul.

# VII. APPENDIX A

## MULTIPLE STEPWISE REGRESSION PROGRAM WITH NEW TERM GENERATING TECHNIQUES

# TABLE OF CONTENTS

# I. INTRODUCTION

A regression analysis provides a relationship between a dependent variable and one or more independent variables. In this program the linear relationship is estimated by least square methods. This program proceeds in a stepwise fashion performing successive regression analyses, including the "best" term first, finding next the term whose additional contribution to the first is greatest, and so on.

The use of a stepwise procedure not only has the advantage of providing useful analytic information to the user, but also provides protection against failures in computational accuracy.

## II. GENERAL DESCRIPTION

This program computes a sequence of multiple linear regression equations in a stepwise manner. Up to fourteen original variables can be used. The variable to be considered as the dependent variable (Y) can be specified on a control card. The remaining variables are considered as the independent variables. If the user would wish not to use all of the independent variables, then a change of format on a control card will allow the use of only the specified independent variables.

Along with each independent variable is associated a maximum power. The maximum power of an independent variable is the highest allowed power to which that independent variable may be raised. The independent variables are then expanded to a new set of terms which includes all possible combinations of the independent variables to varying powers within the maximum power of each independent variable. For example, if three independent variables are used ($Z_1$, $Z_2$, $Z_3$) with a maximum power of 1, 2, and 2 respectively, then the independent terms generated are $X_1=Z_3$, $X_2=Z_3^2$, $X_3=Z_2$, $X_4=Z_2Z_3$, $X_5=Z_2Z_3^2$, $X_6=Z_2^2$, $X_7=Z_2^2Z_3$, $X_8=Z_2^2Z_3^2$, $X_9=Z_1$, $X_{10}=Z_1Z_3$, $X_{11}=Z_1Z_3^2$, $X_{12}=Z_1Z_2$, $X_{13}=Z_1Z_2Z_3$, $X_{14}=Z_1Z_2Z_3^2$, $X_{15}=Z_1Z_2^2$, $X_{16}=Z_1Z_2^2Z_3$, and $X_{17}=Z_1Z_2^2Z_3^2$. Note that the expanded set contains terms (Xs') and not variables (Zs').

The number of independent terms is $\sum_{i=1}^{NM1}(MAXP_i + 1) - 1$, where NM1 is the number of independent variables and $MAXP_i$ is the maximum power of independent variable i.

As can be seen, the list of independent terms could become exorbitant. However, this program includes a system for the user to control the number of terms by specifing the program mode. Three program modes are possible: NORMAL, ACCEPT, and DELETE. If all possible combinations are desired as above, then the NORMAL program mode is used and the terms generated comprise a complete set of terms. However, if only selected terms are desired, then a subset of the complete set of terms can be designated in the ACCEPT program mode. For the above example, the numbers 121 and 101 on a control card would only allow the terms $Z_1Z_2^2Z_3$ and $Z_1Z_3$ to be considered for regression. Moreover, a group of terms that contains all terms with the term $Z_1Z_3$ included in them can be stipulated. For example, 1*1 on a control card would allow the independent terms $Z_1Z_3$, $Z_1Z_2Z_3$, and $Z_1Z_2^2Z_3$ to be generated.

In a similar manner certain terms can be deleted from the complete set of independent terms. This is accomplished by specifing the DELETE program mode. For the above example, if in DELETE mode, the numbers 121 and 101 on a control card would allow all independent terms of the complete set with the exception of $Z_1Z_2^2Z_3$ and $Z_1Z_3$ to be generated. Likewise 1*1 would allow all but the terms $Z_1Z_3$, $Z_1Z_2Z_3$, and $Z_1Z_2^2Z_3$ of the complete set to be generated.

At each step one independent term is added to the regression equation. The term added is the one which makes the greatest reduction in the error sum of squares. Equivalently it is the term which has highest partial correlation with the dependent variable partialed on the independent terms which have already been added; and equivalently it is the term which, if it were added, would have the highest F value. A term may be indicated to be significant in any early stage and thus enter the regression equation. After several other terms are added to the regression equation, the initial term may be indicated to be insignificant. The insignificant term will be removed from the regression equation before adding an additional term. Therefore, only significant terms are included in the final regression equation.

## III. PROGRAM OUTPUT

A. Output prior to performing regression:
1. Problem number
2. F level to enter a term
3. F level to remove a term
4. Number of independent variables
5. Number of independent terms
6. Maximum power of each independent variable
7. Number of observations
8. Weighted number of data
9. Standard error of Y

B. Optional output prior to performing regression:
1. Input data
2. Weighted sum of terms
3. Weighted mean of terms
4. Weighted sum of squares and cross products
5. Weighted residual sum of squares and cross products
6. Correlation coefficients
7. List of term numbers generated (used only when in accept or delete mode)

C. At each step:
1. Step number
2. Term entered or removed
3. F level
4. Standard error of Y
5. Analysis of variance table
6. For terms in the equation
   a. Regression coefficient
   b. Standard error of coefficient
   c. Powers of the independent variables
7. Constant term
8. Multiple correlation coefficient squared

D. Optional output after performing regression:
1. List of residuals

Since the user is not aware whether terms or variables are being used, the word "variable" will appear on the output pages. This is the more commonly accepted terminology.

## IV. PROBLEM LIMITATIONS

1. Number of original variables (14)
2. Number of terms (150)
3. F level to enter a term $\geq$ F level to remove a term
4. Maximum power of an independent variable is nine

## V. ORDER OF CONTROL CARDS IN JOB DECK

1. Start of input data control card
2. Input data cards - place data input deck here
3. End of input data control card
4. Information control card
5. Print control card
6. Maximum powers control card
7. Format control card
8. Variable mode control card
9. Variable mode control data cards

If more than one analysis of a set of input data is required, cards 4-9 can be repeated. If more than one set of input data needs to be analyzed, repeat cards 1-9.

## VI. CONTROL CARD PREPARATION

1. Start of input data control card

| | | |
|---|---|---|
| Col. | 1 - 10 | START DATA |

3. End of input data control card

| | | |
|---|---|---|
| Col. | 1 - 8 | END DATA |

4. Information control card

| | | |
|---|---|---|
| Col. | 1 - 6 | INFORM |
| Col. | 11-15 | Numeric problem number |
| Col. | 16-20 | Number of original variables |
| Col. | 21-25 | F level to enter a term (X. XX) |
| Col. | 26-30 | F level to remove a term (X. XX) |
| Col. | 35 | 1 if all observations are equally weighted; otherwise, 0 |

5. Print card

| | | |
|---|---|---|
| Col. | 1 - 5 | PRINT |
| Col. | 15 | 1 if the input data is to be printed; otherwise, 0 |
| Col. | 20 | 1 if the weighted sums of terms are to be printed; otherwise, 0 |
| Col. | 25 | 1 if the weighted means of terms are to be printed; otherwise, 0 |
| Col. | 30 | 1 if the weighted sums of squares and cross products are to be printed; otherwise, 0 |
| Col. | 35 | 1 if the weighted residual sums of squares and cross products are to be printed; otherwise, 0 |
| Col. | 40 | 1 if the correlation coefficients are to be printed; otherwise, 0 |
| Col. | 45 | 1 if the list that transforms term numbers from ACCEPT or DELETE mode to NORMAL mode is to be printed; otherwise, 0 (will always be 0 in the NORMAL mode) |
| Col. | 50 | 1 if the residuals are to be printed; otherwise, 0 |

6. Maximum powers control card

| | | |
|---|---|---|
| Col. | 1 - 10 | MAX POWERS |
| Col. | 10+5I | Maximum power allowed for independent variable I |

7. Format control card

| | |
|---|---|
| Col. 1 - 6 | FORMAT |
| Col. 7 -78 | Format of the input data enclosed in parenthesis |
| Col. 79-80 | Number of the variable designated as the dependent variable |

8. Variable mode control card

| | |
|---|---|
| Col. 1 - 6 | NORMAL, DELETE, or ACCEPT |
| Col. 11-15 | Number of variable mode control data cards (if in the NORMAL mode, this number will be 0) |

9. Variable mode control data cards

These cards contain the powers of the independent variables used to compose each term of the subset of independent terms to be accepted or deleted from the complete set of terms. An asterisk in place of a power allows all powers up to its maximum power of that particular independent variable to be accepted or deleted. For each term (or group of terms if asterisks are used) the number of powers and asterisks must equal the number of independent variables used. Blanks can be placed between powers and asterisks. A comma must separate the powers of one term from the powers of another term.

# VII. EXAMPLES OF CONTROL CARDS

PROGRAM: MULTIPLE STEPWISE REGRESSION ANALYSIS
PROGRAMMER: RICHARD R. KUNKEL
DATE: 1970
PUNCHING INSTRUCTIONS — GRAPHIC / PUNCH
PAGE 1 OF 1
CARD ELECTRO NUMBER

STATEMENT (Name 1–8, Operation 10–14, Operand 16–, Comments; Identification-Sequence 73–80)

```
START DATA
        7.       26.        1.       78.       6.5       60.
       10.       68.       13.     109.4        8.       12.
END DATA
INFORM        1    5 2.50 2.50    1
PRINT         0    0    1    0    0    1    0    1
MAX POWERS    3    3    2    2
FORMAT(2F10.3,10X,3F10.3)                                                      3
ACCEPT        2
  1200,  0     0   1     2 ,00    02                         ,    3
2   1  0  ,           0                 0        2                       1
INFORM        2    6 2.00 2.00    1
PRINT         0    0    0    0    0    0    0    0
MAX POWERS    4    1    5    2    3
FORMAT(6F10.3)                                                                 4
ACCEPT        1
   **511,11  11 1,   4 14*2
```

Two runs on the same set of data have been indicated.

Run one:

Problem number 1

Five original variables

F level to enter a term is 2.50

F level to remove a term is 2.50

All observations are equally weighted

Weighted means of terms, correlation coefficients, and residuals will be printed

There are four independent variables ($Z_1$, $Z_2$, $Z_3$, $Z_4$) with highest power allowed of 3, 2, 2, 2 respectively

The first ten columns of each data card contain the value of $Z_1$, columns 11-20 the value of $Z_2$, columns 31-40 the value of the dependent variable (Y), columns 41-50 the value of $Z_3$, and columns 51-60 the value of $Z_4$

The program is in the ACCEPT mode and has two variable mode control data cards

The independent terms to be considered for regression are $Z_1Z_2^2$, $Z_3Z_4^2$, $Z_4^2$, $Z_1^3Z_2^2Z_3$, and $Z_3^2Z_4$

Run two:

Problem number 2

Six original variables

F level to enter a term is 2.00

F level to remove a term is 2.00

All observations are equally weighted

No optional information is to be printed

There are five independent variables ($Z_1$, $Z_2$, $Z_3$, $Z_4$, $Z_5$) with highest power allowed of 4, 1, 5, 2, 3 respectively

Starting with column one on each data card, the value of variables $Z_1$, $Z_2$, $Z_3$, Y, $Z_4$, and $Z_5$ are read from consecutive fields of ten columns each

The program is in the ACCEPT mode and has one variable mode control data card

The independent terms to be considered for regression are $Z_3^5Z_4Z_5$, $Z_2Z_3^5Z_4Z_5$, $Z_1Z_3^5Z_4Z_5$, $Z_1Z_2Z_3^5Z_4Z_5$, $Z_1^2Z_3^5Z_4Z_5$, $Z_1^2Z_2Z_3^5Z_4Z_5$, $Z_1^3Z_3^5Z_4Z_5$, $Z_1^3Z_2Z_3^5Z_4Z_5$, $Z_1^4Z_3^5Z_4Z_5$, $Z_1^4Z_2Z_3^5Z_4Z_5$, $Z_1Z_2Z_3Z_4Z_5$, $Z_1^4Z_2Z_3^4Z_5^2$, $Z_1^4Z_2Z_3^4Z_4Z_5^2$, and $Z_1^4Z_2Z_3^4Z_4^2Z_5^2$

# VIII. BLOCK FLOW CHART

START

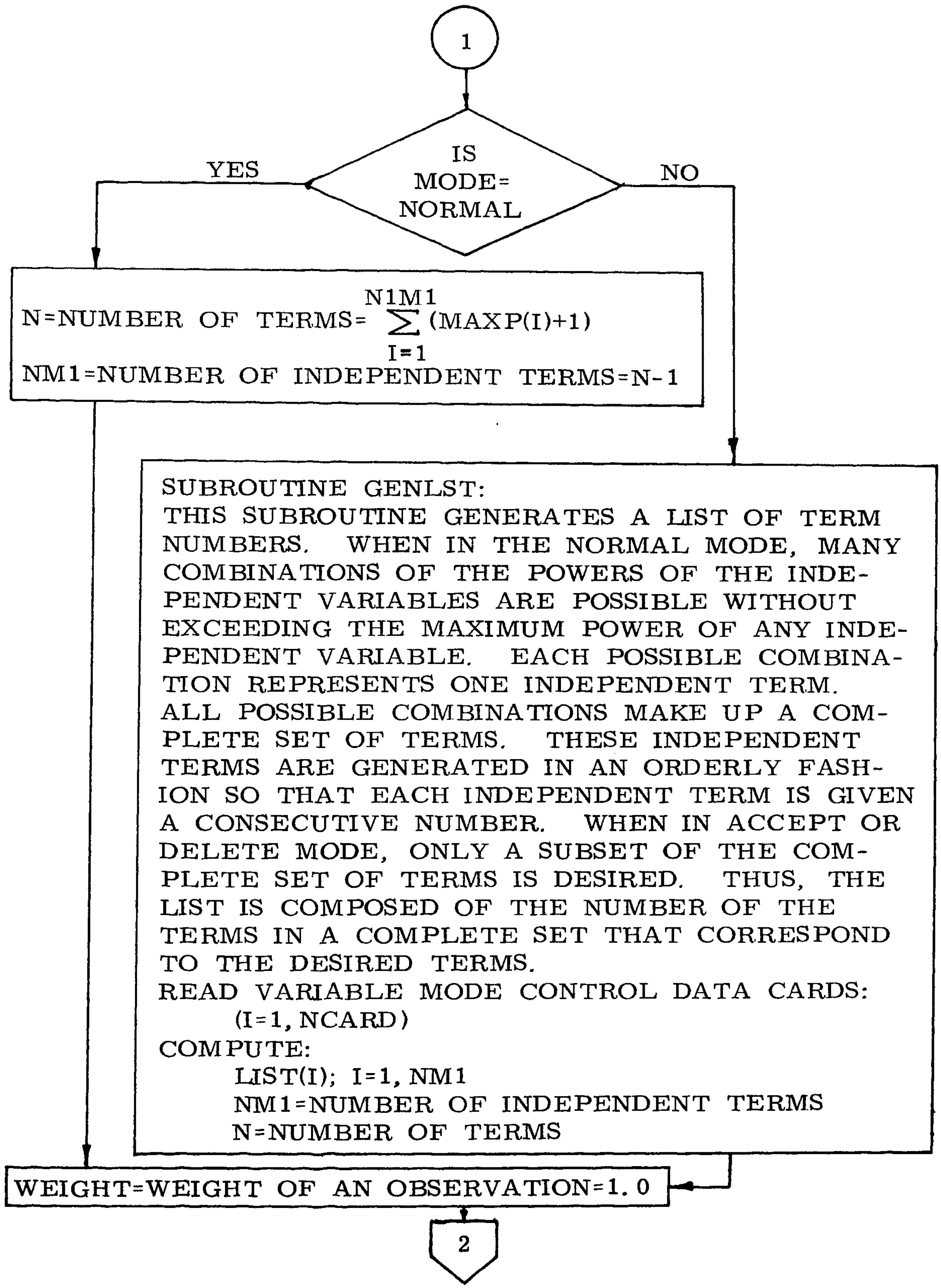
READ INFORMATION CONTROL CARD:
N1=NO. OF INDEPENDENT VARIABLES + ONE DEPENDENT VARIABLE
F1=F VALUE TO ENTER A VARIABLE
F2=F VALUE TO REMOVE A VARIABLE

READ IN ONE INPUT DATA CARD AND WRITE THE CARD IMAGE ON A SPECIFIED UNIT. CONTINUE THIS PROCESS UNTIL ALL INPUT DATA CARDS HAVE BEEN READ

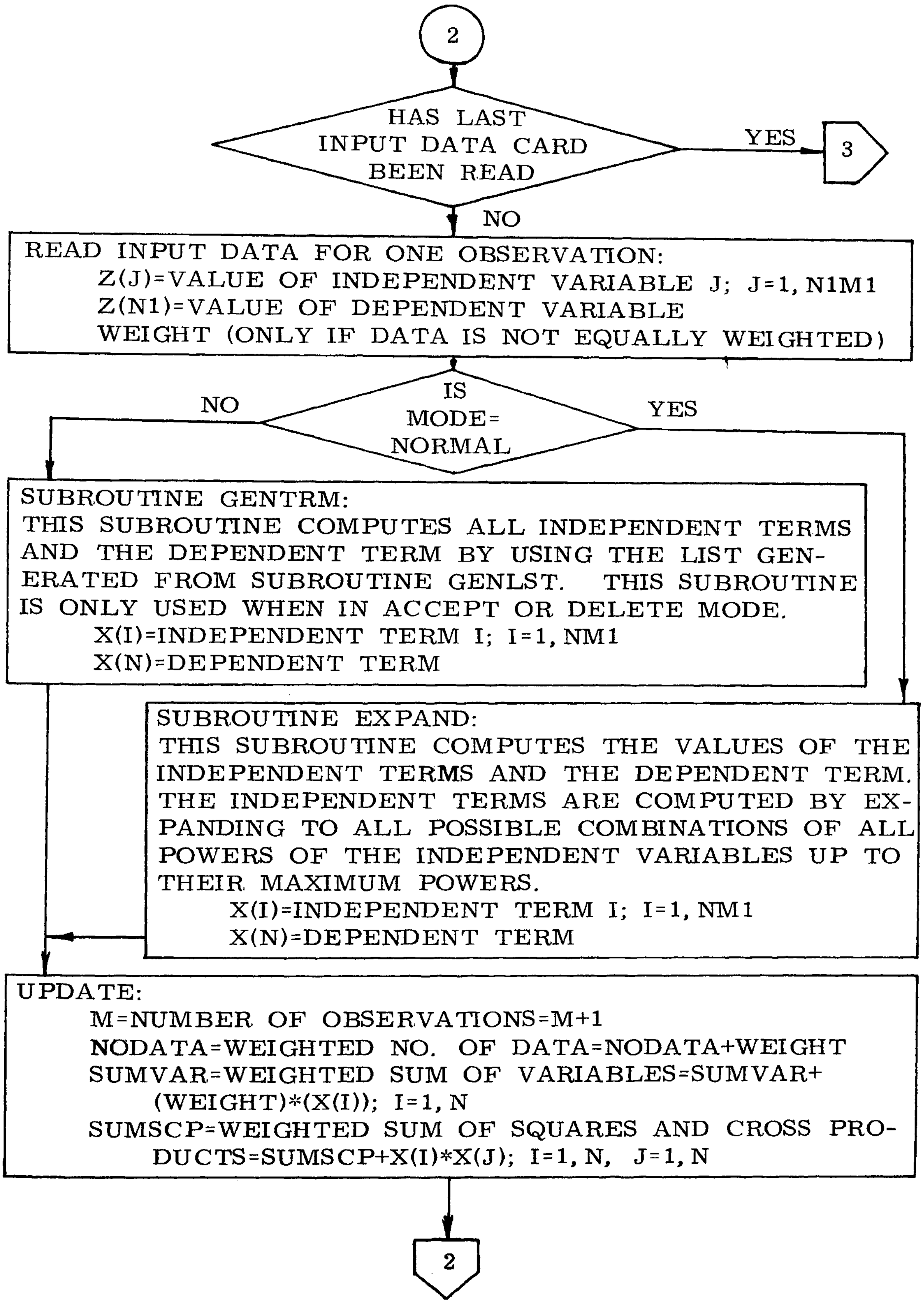
TOL=TOLERANCE TO CONTROL DEGENERACY=.001
N1M1=NUMBER OF INDEPENDENT VARIABLES=N1-1

READ MAXIMUM POWERS CONTROL CARD:
MAXP(I)=MAXIMUM POWER OF INDEPENDENT VARIABLE I; I=1, N1M1
READ FORMAT CONTROL CARD:
FMT=ARRAY THAT CONTAINS ALPHA INFORMATION ABOUT THE FORMAT OF THE INPUT DATA
READ VARIABLE MODE CONTROL CARD:
MODE=VARIABLE MODE (ACCEPT, DELETE, OR NORMAL)
NCARD=NUMBER OF VARIABLE MODE CONTROL DATA CARDS

1



1
IS MODE= NORMAL
YES
NO
N=NUMBER OF TERMS= $\sum_{I=1}^{N1M1}$ (MAXP(I)+1)
NM1=NUMBER OF INDEPENDENT TERMS=N-1
SUBROUTINE GENLST:
THIS SUBROUTINE GENERATES A LIST OF TERM NUMBERS. WHEN IN THE NORMAL MODE, MANY COMBINATIONS OF THE POWERS OF THE INDEPENDENT VARIABLES ARE POSSIBLE WITHOUT EXCEEDING THE MAXIMUM POWER OF ANY INDEPENDENT VARIABLE. EACH POSSIBLE COMBINATION REPRESENTS ONE INDEPENDENT TERM. ALL POSSIBLE COMBINATIONS MAKE UP A COMPLETE SET OF TERMS. THESE INDEPENDENT TERMS ARE GENERATED IN AN ORDERLY FASHION SO THAT EACH INDEPENDENT TERM IS GIVEN A CONSECUTIVE NUMBER. WHEN IN ACCEPT OR DELETE MODE, ONLY A SUBSET OF THE COMPLETE SET OF TERMS IS DESIRED. THUS, THE LIST IS COMPOSED OF THE NUMBER OF THE TERMS IN A COMPLETE SET THAT CORRESPOND TO THE DESIRED TERMS.
READ VARIABLE MODE CONTROL DATA CARDS:
(I=1, NCARD)
COMPUTE:
LIST(I); I=1, NM1
NM1=NUMBER OF INDEPENDENT TERMS
N=NUMBER OF TERMS
WEIGHT=WEIGHT OF AN OBSERVATION=1.0
2


2
HAS LAST INPUT DATA CARD BEEN READ
YES
3
NO
READ INPUT DATA FOR ONE OBSERVATION:
Z(J)=VALUE OF INDEPENDENT VARIABLE J; J=1, N1M1
Z(N1)=VALUE OF DEPENDENT VARIABLE
WEIGHT (ONLY IF DATA IS NOT EQUALLY WEIGHTED)
NO
IS MODE= NORMAL
YES
SUBROUTINE GENTRM:
THIS SUBROUTINE COMPUTES ALL INDEPENDENT TERMS AND THE DEPENDENT TERM BY USING THE LIST GENERATED FROM SUBROUTINE GENLST. THIS SUBROUTINE IS ONLY USED WHEN IN ACCEPT OR DELETE MODE.
X(I)=INDEPENDENT TERM I; I=1, NM1
X(N)=DEPENDENT TERM
SUBROUTINE EXPAND:
THIS SUBROUTINE COMPUTES THE VALUES OF THE INDEPENDENT TERMS AND THE DEPENDENT TERM. THE INDEPENDENT TERMS ARE COMPUTED BY EXPANDING TO ALL POSSIBLE COMBINATIONS OF ALL POWERS OF THE INDEPENDENT VARIABLES UP TO THEIR MAXIMUM POWERS.
X(I)=INDEPENDENT TERM I; I=1, NM1
X(N)=DEPENDENT TERM
UPDATE:
M=NUMBER OF OBSERVATIONS=M+1
NODATA=WEIGHTED NO. OF DATA=NODATA+WEIGHT
SUMVAR=WEIGHTED SUM OF VARIABLES=SUMVAR+(WEIGHT)*(X(I)); I=1, N
SUMSCP=WEIGHTED SUM OF SQUARES AND CROSS PRODUCTS=SUMSCP+X(I)*X(J); I=1, N, J=1, N
2

3

COMPUTE:

XBAR(I)=WEIGHTED MEAN= $\sum_{t=1}^{M} WEIGHT_t \; X_{it} / \sum_{t=1}^{M} WEIGHT_t$

=SUMVAR(I)/NODATA; I=1, N

S(I, J)=WT. RESIDUAL SUM OF SQUARES & CROSS PRODUCTS

$= \left( \sum_{t=1}^{M} WT_t \sum_{t=1}^{M} WT_t \; X_{it} \; X_{jt} - \sum_{t=1}^{M} WT_t \; X_{it} \sum_{t=1}^{M} WT_t \; X_{jt} \right) / \sum_{t=1}^{M} WT_t$

=SUMSCP(I, J)-XBAR(I)*XBAR(J)*NODATA; I=1, N, J=I, N

SIGMA(I)=SQRT(S(I, I)); I=1, N

R(I, J)=SIMPLE CORRELATION COEFFICIENTS

=S(I, J)/SIGMA(I)*SIGMA(J); I=1, NM1, J=I+1, N

R(J, I)=R(I, J)

R(I, I)=1.00; I=1, N

DF=DEGREES OF FREEDOM= $\sum_{t=1}^{M} WEIGHT_t - 1.0$ =NODATA-1.0

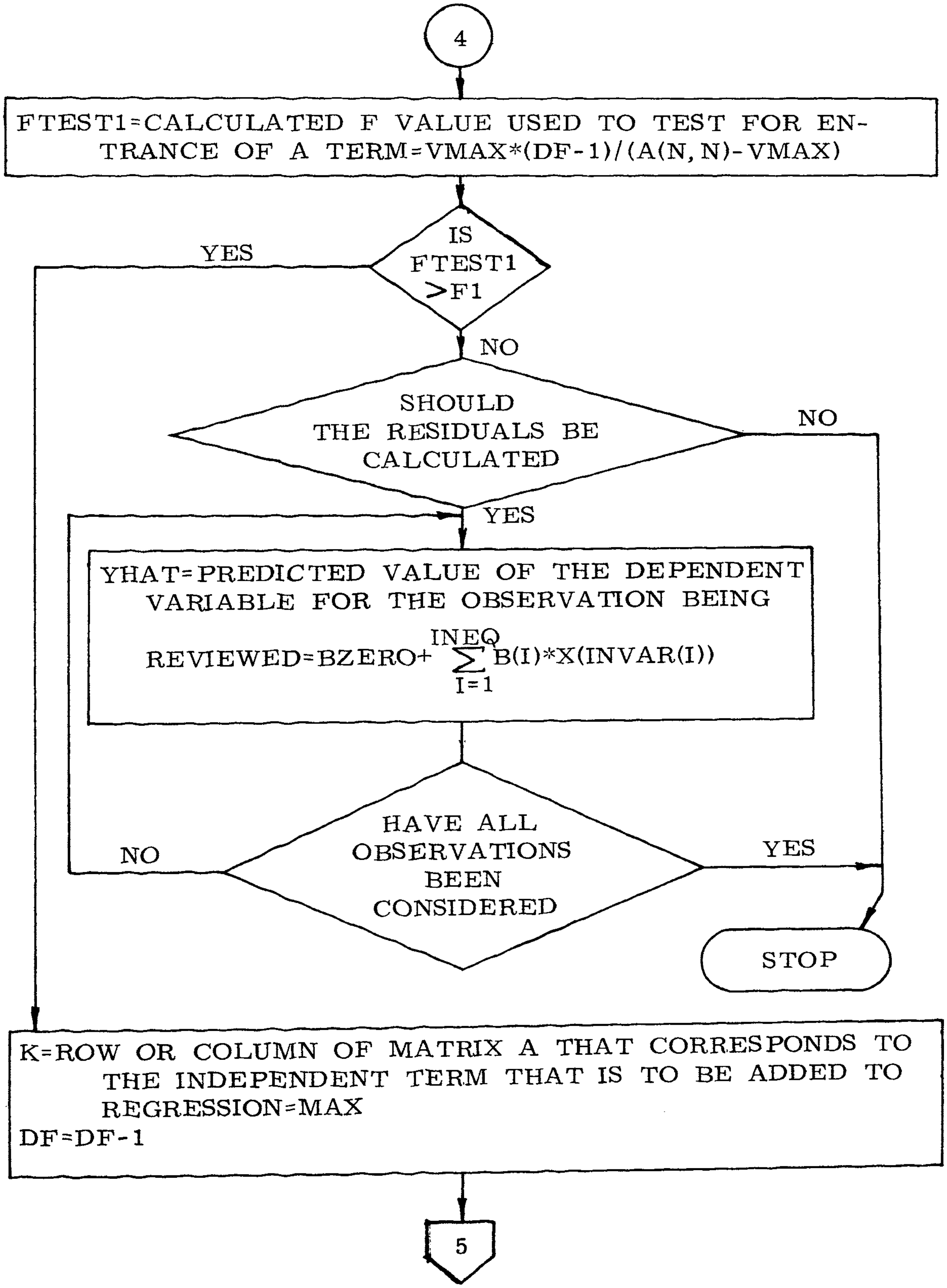
ERRORY=STANDARD ERROR OF Y=SIGMA(N)*SQRT(R(N, N)/DF)

MAX=VALUE OF I CORRESPONDING TO LARGEST ABSOLUTE VALUE OF R(I, N); I=1, NM1

A(I, J)=R(I, J); I=1, N, J=1, N

VMAX=VARIANCE REDUCTION BY ADDING VARIABLE MAX TO REGRESSION=A(MAX, N)*A(N, MAX)

4


4
FTEST1=CALCULATED F VALUE USED TO TEST FOR ENTRANCE OF A TERM=VMAX*(DF-1)/(A(N,N)-VMAX)
IS FTEST1 >F1
YES
NO
SHOULD THE RESIDUALS BE CALCULATED
NO
YES
YHAT=PREDICTED VALUE OF THE DEPENDENT VARIABLE FOR THE OBSERVATION BEING REVIEWED=BZERO+ $\sum_{I=1}^{INEQ}$ B(I)*X(INVAR(I))
HAVE ALL OBSERVATIONS BEEN CONSIDERED
NO
YES
STOP
K=ROW OR COLUMN OF MATRIX A THAT CORRESPONDS TO THE INDEPENDENT TERM THAT IS TO BE ADDED TO REGRESSION=MAX
DF=DF-1
5

5

CALCULATE NEXT MATRIX:

1. $A(I, J)_{new} = (A(K, K)*A(I, J) - A(I, K)*A(K, J))/A(K, K)$; $I \neq K$, $J \neq K$

2. $A(I, J)_{new} = -A(I, K)/A(K, K)$; $I \neq K$, $J = K$

3. $A(I, J)_{new} = A(K, J)/A(K, K)$; $J \neq K$, $I = K$

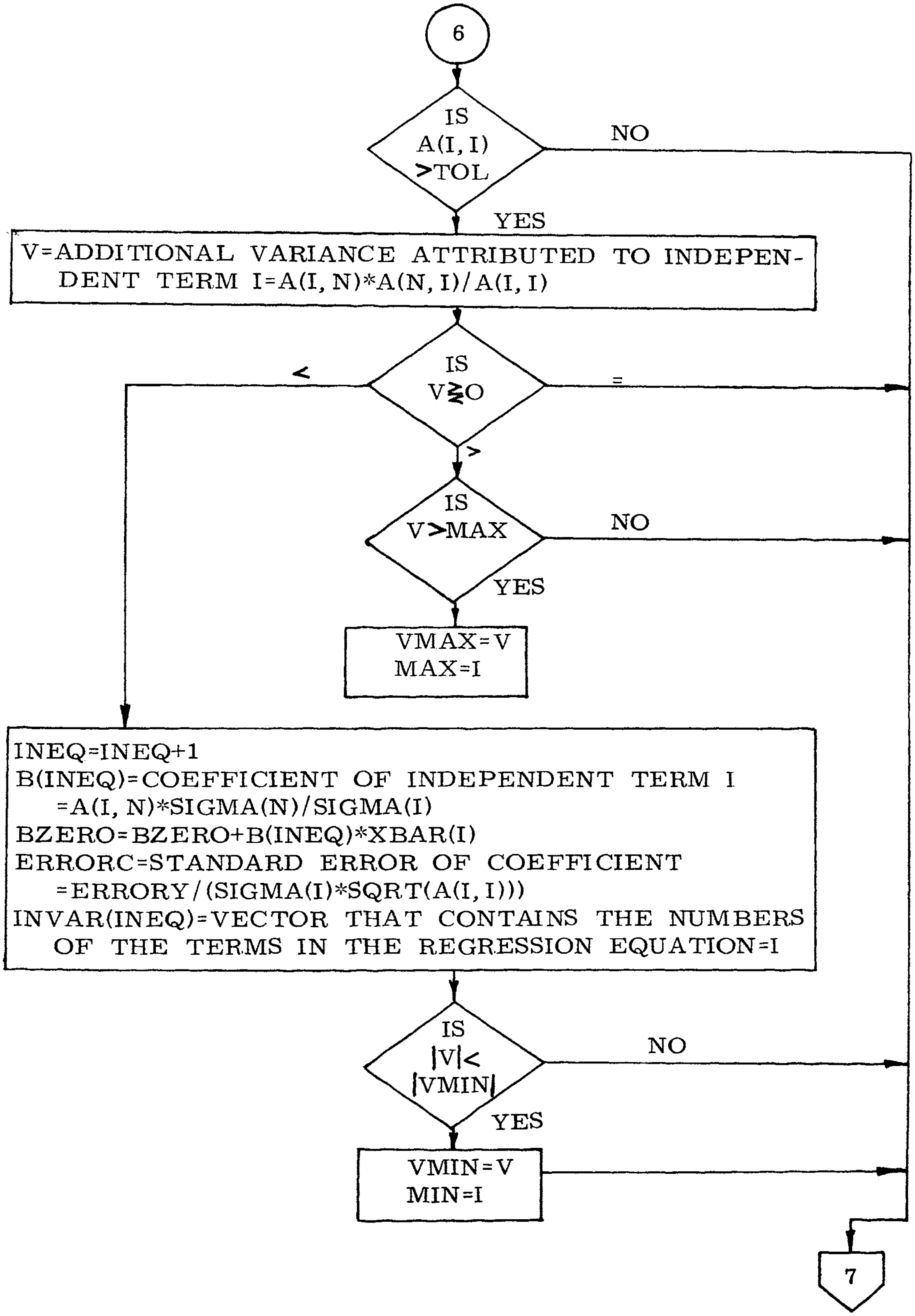
4. $A(K, K)_{new} = 1/A(K, K)$

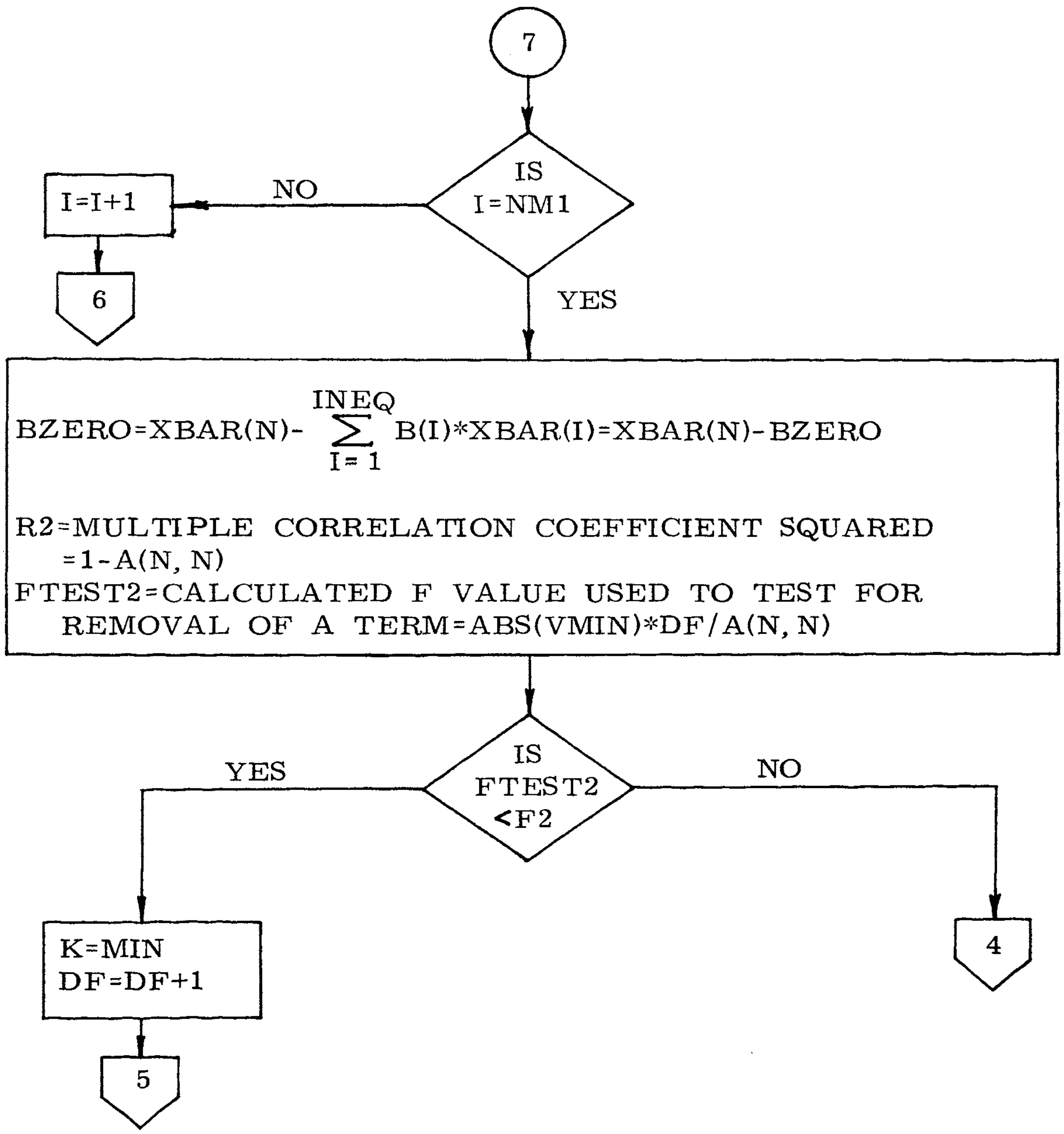
ERRORY=SIGMA(N)*SQRT(A(N, N)/DF)

COMPUTE PARAMETERS FOR AOV TABLE:
TOALSS=TOTAL SUM OF SQUARES=SUMSCP(N, N)
MEANSS=MEAN SUM OF SQUARES=TOALSS-S(N, N)
REGDF=DEGREES OF FREEDOM ATTRIBUTED TO
REGRESSION=NODATA-DF-1
RESSS=RESIDUAL SUM OF SQUARES
=A(N, N)*(TOALSS-MEANSS)
REGSS=REGRESSION SUM OF SQUARES
=TOALSS-MEANSS-RESSS
REGMS=REGRESSION MEAN SQUARE=REGSS/REGDF
RESMS=RESIDUAL MEAN SQUARE=RESSS/DF
RATIO=F RATIO=REGMS/RESMS

INITIALIZATION:
I=VARIABLE NUMBER=1
INEQ=COUNTER OF THE NUMBER OF TERMS IN
REGRESSION=0
VMIN=VARIANCE INCREASE BY REMOVING A VARIABLE
=5. E10
VMAX=VARIANCE REDUCTION BY ADDING A VARIABLE
=0. 0
MIN=NUMBER OF THE IND. TERM THAT CAUSES THE
LEAST VARIANCE INCREASE WHEN TAKEN OUT OF
REGRESSION=0
MAX=NUMBER OF THE IND. TERM THAT CAUSES THE
GREATEST VARIANCE REDUCTION WHEN PUT INTO
REGRESSION=0
BZERO=CONSTANT TERM=0

6


6
IS
A(I, I)
>TOL
NO
YES
V=ADDITIONAL VARIANCE ATTRIBUTED TO INDEPEN-
DENT TERM I=A(I, N)*A(N, I)/A(I, I)
IS
V≷O
<
=
>
IS
V>MAX
NO
YES
VMAX=V
MAX=I
INEQ=INEQ+1
B(INEQ)=COEFFICIENT OF INDEPENDENT TERM I
=A(I, N)*SIGMA(N)/SIGMA(I)
BZERO=BZERO+B(INEQ)*XBAR(I)
ERRORC=STANDARD ERROR OF COEFFICIENT
=ERRORY/(SIGMA(I)*SQRT(A(I, I)))
INVAR(INEQ)=VECTOR THAT CONTAINS THE NUMBERS
OF THE TERMS IN THE REGRESSION EQUATION=I
IS
|V|<
|VMIN|
NO
YES
VMIN=V
MIN=I
7


7
IS
I=NM1
NO
I=I+1
6
YES
BZERO=XBAR(N)- $\sum_{I=1}^{INEQ}$ B(I)*XBAR(I)=XBAR(N)-BZERO
R2=MULTIPLE CORRELATION COEFFICIENT SQUARED
=1-A(N, N)
FTEST2=CALCULATED F VALUE USED TO TEST FOR
REMOVAL OF A TERM=ABS(VMIN)*DF/A(N, N)
IS
FTEST2
<F2
YES
NO
K=MIN
DF=DF+1
5
4

# IX. PROGRAM LISTING

```
C      MULTIPLE REGRESSION ANALYSIS (STEPWISE)
       REAL Z(15),X(150),SUMVAR(150),XBAR(150),SUMSCP(150, 150),S(150, 150)
      1,SIGMA(150), R(150, 150), A(150, 150), NODATA, MEANSS, B(150)
       INTEGER DF,STEPNO, PRONUM, MAXP(15), P(15), REGDF, LIST(150), CONTRL,
      1DEPVAR, FMT(18), INVAR(150), INPTAP(20)
       DOUBLE PRECISION LEFT3, LEFT5
       DATA LEFT1/2HX-/, LEFT2/2HZ(/, LEFT3/6HSUM X(/, LEFT4/2HX(/,
      1LEFT5/5H * Z(/, IFORM/4HFORM/, ISTAR/4HSTAR/, IEND/4HEND /,
      2INFO/4HINFO/, IPRIN/4HPRIN/, IMAX/4HMAX /, INORM/4HNORM/,
      3IACCE/4HACCE/, IDELE/4HDELE/
       EQUIVALENCE (SUMVAR, XBAR), (SUMSCP, S, R, A), (X, SIGMA)
       TOL=.001
    10 READ(1, 1000, END=620)CONTRL, PRONUM, N1, F1, F2, IW
  1000 FORMAT(A4, 6X, 2I5, 2F5. 2, I5)
    20 IF(CONTRL. NE. ISTAR) GO TO 50
    30 READ(1, 1010)INPTAP
  1010 FORMAT(20A4)
       IF(INPTAP(1). EQ. IEND) GO TO 40
       WRITE(4, 1010)INPTAP
       GO TO 30
    40 ENDFILE 4
       REWIND 4
       GO TO 10
    50 IF(CONTRL. EQ. INFO) GO TO 60
       WRITE(3, 3000)
  3000 FORMAT(' INCORRECT OR MISSING CONTROL CARD - EITHER INFORMATION CA
      1RD OR START OF INPUT DATA CARD')
       GO TO 620
    60 WRITE(3, 3010)PRONUM, N1, F1, F2
```

```
3010 FORMAT('1PROBLEM NO',I5//' NO OF VARIABLES =',I5//' F LEVEL TO ENT
    1ER VARIABLE =',F5.2//' F LEVEL TO REMOVE VARIABLE =',F5.2)
     READ(1,1020)CONTRL,INDATA,ISUMVR,IXBAR,ISMSCP,IS,IR,ILIST,IRESID
1020 FORMAT(A4,6X,14I5)
     IF(CONTRL.EQ.IPRIN) GO TO 70
     WRITE(3,3020)
3020 FORMAT(' INCORRECT OR MISSING CONTROL CARD - PRINT CARD')
     GO TO 620
  70 N1M1=N1-1
     READ(1,1020)CONTRL,(MAXP(I),I=1,N1M1)
     IF(CONTRL.EQ.IMAX) GO TO 80
     WRITE(3,3030)
3030 FORMAT(' INCORRECT OR MISSING CONTROL CARD - MAXIMUM POWERS CARD')
     GO TO 620
  80 READ(1,1030)CONTRL,FMT,DEPVAR
1030 FORMAT(A4,2X,18A4,I2)
     IF(CONTRL.EQ.IFORM) GO TO 90
     WRITE(3,3040)
3040 FORMAT(' INCORRECT OR MISSING CONTROL CARD - FORMAT CARD')
     GO TO 620
  90 READ(1,1020)MODE,NCARD
     IF(MODE.EQ.INORM) GO TO 110
     IF((MODE.EQ.IDELE).OR.(MODE.EQ.IACCE)) GO TO 100
     WRITE(3,3050)
3050 FORMAT(' INCORRECT OR MISSING CONTROL CARD - VARIABLE MODE CARD')
     GO TO 620
 100 CALL GENLST(MODE,N1M1,N,NM1,NCARD,MAXP(1),LIST(1),A(1,1),A(1,76))
     GO TO 130
 110 N=1
     DO 120 I=1,N1M1
 120 N=N*(MAXP(I)+1)
```

```
      NM1=N-1
  130 WRITE(3,3060)N1M1,NM1,(LEFT1,I,MAXP(I),I=1,N1M1)
 3060 FORMAT(/' NO. OF INDEPENDENT VARIABLES =',I3//' NO. OF INDEPENDENT
     1 TERMS =',I4//3X,'INDEPENDENT',4X,'MAX'/5X,'VARIABLE',4X,'POWER'/
     2(7X,A2,I2,7X,I2))
      IF(INDATA.EQ.0) GO TO 140
      WRITE(3,3070)(LEFT2,(I),I=1,N1)
 3070 FORMAT(///20X,'INPUT DATA'//' OBSERVATION',10(3X,A2,I2,')',2X)/
     1(15X,A2,I2,')',5X,A2,I2,')',5X,A2,I2,')',5X,A2,I2,')',5X,A2,I2,')'
     2,5X,A2,I2,')',5X,A2,I2,')',5X,A2,I2,')',5X,A2,I2,')',5X,A2,I2,')')
     3)
  140 DO 150 I=1,N
      SUMVAR(I)=0.0
      DO 150 J=1,N
  150 SUMSCP(I,J)=0.0
      NODATA=0.0
      WEIGHT=1.0
      M=0
  160 IF(IW.EQ.0) GO TO 162
      READ(4,FMT,END=240)(Z(J),J=1,N1)
      GO TO 170
  162 READ(4,FMT,END=240)(Z(J),J=1,N1),WEIGHT
  170 M=M+1
      TEMP=Z(DEPVAR)
      DO 180 J=DEPVAR,N1M1
  180 Z(J)=Z(J+1)
      Z(N1)=TEMP
      IF(INDATA.EQ.0) GO TO 190
      WRITE(3,3080)M,(Z(J),J=1,N1)
 3080 FORMAT(3X,I4,10F10.3,(/11X,10F10.3))
  190 NODATA=NODATA+WEIGHT
```

```
      IF(MODE. EQ. INORM) GO TO 200
      CALL GENTRM(N1M1, NM1, MAXP(1), LIST(1), Z(1), X(1))
      GO TO 210
  200 CALL EXPAND(N1M1, MAXP(1), Z(1), X(1))
  210 DO 220 J=1, N
  220 SUMVAR(J)=SUMVAR(J)+WEIGHT*X(J)
      DO 230 I=1, N
      DO 230 J=I, N
  230 SUMSCP(I, J)=SUMSCP(I, J)+WEIGHT*X(I)*X(J)
      GO TO 160
  240 WRITE(3, 3090)M, NODATA
 3090 FORMAT(/' NO. OF OBSERVATIONS =', I5//' WEIGHTED NO. OF DATA =',
     1F7. 2)
      IF(ISUMVR. EQ. 0) GO TO 250
      WRITE(3, 3100)
 3100 FORMAT(///20X, 'WEIGHTED SUMS OF VARIABLES'/)
      WRITE(3, 3110)((LEFT3, I, SUMVAR(I)), I=1, NM1)
 3110 FORMAT(3(1X, A6, I2, ') =', F14. 5, 4X))
      WRITE(3, 3120)SUMVAR(N)
 3120 FORMAT(' SUM    Y    =', F14. 5)
  250 DO 260 I=1, N
  260 XBAR(I)=SUMVAR(I)/NODATA
      IF(IXBAR. EQ. 0) GO TO 270
      WRITE(3, 3130)
 3130 FORMAT(///20X, 'WEIGHTED MEAN OF VARIABLES'/)
      WRITE(3, 3110)((LEFT3, I, XBAR(I)), I=1, NM1)
      WRITE(3, 3120)XBAR(N)
  270 IF(ISMSCP. EQ. 0) GO TO 280
      WRITE(3, 3140)
 3140 FORMAT(///20X, 'WEIGHTED SUMS OF SQUARES AND CROSS PRODUCTS'/)
      WRITE(3, 3150)(((LEFT4, I, J, SUMSCP(I, J)), J=I, NM1), I=1, NM1)
```

```
3150 FORMAT(3(1X, A2, I2, ') VS X(', I2, ') =', F16.5, 4X))
     WRITE(3, 3160)((LEFT4, I, SUMSCP(I, N)), I=1, NM1)
3160 FORMAT(3(1X, A2, I2, ') VS    Y    =', F16.5, 4X))
     WRITE(3, 3170)SUMSCP(N, N)
3170 FORMAT(' Y      VS    Y    =', F16.5)
 280 TOALSS=SUMSCP(N, N)
     DO 300 I=1, N
     DO 290 J=I, N
 290 S(I, J)=SUMSCP(I, J)-XBAR(I)*XBAR(J)*NODATA
 300 SIGMA(I)=SQRT(S(I, I))
     IF(IS.EQ.0) GO TO 310
     WRITE(3, 3180)
3180 FORMAT(///20X, 'WEIGHTED RESIDUAL SUM OF SQUARES AND CROSS PRODUCTS
    1'/)
     WRITE(3, 3150)(((LEFT4, I, J, S(I, J)), J=I, NM1), I=1, NM1)
     WRITE(3, 3160)((LEFT4, I, S(I, N)), I=1, NM1)
     WRITE(3, 3170)S(N, N)
 310 MEANSS=TOALSS-S(N, N)
     DO 320 I=1, NM1
     IP1=I+1
     DO 320 J=IP1, N
 320 R(I, J)=S(I, J)/(SIGMA(I)*SIGMA(J))
     DO 330 I=1, NM1
     IP1=I+1
     DO 330 J=IP1, N
 330 R(J, I)=R(I, J)
     DO 340 I=1, N
 340 R(I, I)=1.0
     IF(IR.EQ.0) GO TO 350
     WRITE(3, 3190)
3190 FORMAT(///20X, 'CORRELATION COEFFICIENTS'/)
```

```
      WRITE(3,3150)(((LEFT4,I,J,R(I,J)),J=I,NM1),I=1,NM1)
      WRITE(3,3160)((LEFT4,I,R(I,N)),I=1,NM1)
  350 DF=NODATA-1.0
      ERRORY=SIGMA(N)*SQRT(R(N,N)/DF)
      WRITE(3,3200)ERRORY
 3200 FORMAT(/' STANDARD ERROR OF Y =',F8.3)
      IF(ILIST.EQ.0) GO TO 360
      WRITE(3,3210)(I,LIST(I),I=1,NM1)
 3210 FORMAT(/' LIST GENERATION'/3X,'(TERM GENERATED) CORRESPONDS TO THE
     1 X TERM GENERATED IN THE NORMAL MODE'//5X,'TERM NO.',3X,'TERM GENE
     2RATED'/(6X,I5,6X,I5))
  360 VMAX=0.0
      DO 370 I=1,NM1
      IF(ABS(R(I,N)).LE.VMAX) GO TO 370
      MAX=I
      VMAX=ABS(R(I,N))
  370 CONTINUE
      VMAX=A(MAX,N)*A(N,MAX)
      STEPNO=1
      GO TO 480
  380 STEPNO=STEPNO+1
      ERRORY=SIGMA(N)*SQRT(A(N,N)/DF)
      WRITE(3,3220)ERRORY
 3220 FORMAT(3X,'STANDARD ERROR OF Y =',F12.4/3X,'ANALYSIS OF VARIANCE'
     1/12X,'SOURCE',6X,'DF',4X,'SUM OF SQUARES',4X,'MEAN SQUARE',3X,
     2'F RATIO')
      WRITE(3,3230)M,TOALSS
 3230 FORMAT(10X,'TOTAL',7X,I4,4X,F14.3)
      WRITE(3,3240)MEANSS
 3240 FORMAT(10X,'MEAN',11X,'1',4X,F14.3)
      REGDF=NODATA-DF-1
```

```
      RESSS=A(N, N)*(T)ALSS-MEANSS)
      REGSS=TOALSS-MEANSS-RESSS
      REGMS=REGSS/REGDF
      RESMS=RESSS/DF
      RATIO=REGMS/RESMS
      WRITE(3, 3250)REGDF, REGSS, REGMS, RATIO
 3250 FORMAT(10X, 'REGRESSION', 2X, I4, 4X, F14. 3, 2X, F13. 3, 2X, F8. 3)
      WRITE(3, 3260)DF, RESSS, RESMS
 3260 FORMAT(10X, 'RESIDUAL', 4X, I4, 4X, F14. 3, 2X, F13. 3/3X, 'VARIABLES IN EQU
     1ATION'/10X, 'VARIABLE', 5X, 'COEFFICIENT', 5X, 'STD ERROR OF COEF')
      I=1
      INEQ=0
      VMIN=5. E10
      VMAX=0. 0
      MIN=0
      MAX=0
      BZERO=0. 0
  390 IF(A(I, I). LE. TOL)GO TO 420
      V=A(I, N)*A(N, I)/A(I, I)
      IF(V)410, 420, 400
  400 IF(V. LE. VMAX)GO TO 420
      VMAX=V
      MAX=I
      GO TO 420
  410 INEQ=INEQ+1
      B(INEQ)=A(I, N)*SIGMA(N)/SIGMA(I)
      BZERO=BZERO+B(INEQ)*XBAR(I)
      ERRORC=ERRORY/SIGMA(I)*SQRT(A(I, I))
      WRITE(3, 3270)I, B(INEQ), ERRORC
 3270 FORMAT(12X, 'X-', I3, 6X, F10. 5, 7X, F10. 5)
      INVAR(INEQ)=I
```

```
      IF(ABS(V).GE.ABS(VMIN)) GO TO 420
      VMIN=V
      MIN=I
  420 IF(I.EQ.NM1) GO TO 430
      I=I+1
      GO TO 390
  430 DO 470 J=1,INEQ
      IF(MODE.EQ.INORM) GO TO 440
      IRET=LIST(INVAR(J))
      GO TO 450
  440 IRET=INVAR(J)
  450 DO 460 IJ=1,N1M1
      IZ=IRET/(MAXP(N1-IJ)+1)
      P(N1-IJ)=IRET-IZ*(MAXP(N1-IJ)+1)
  460 IRET=IZ
  470 WRITE(3,3280)INVAR(J),P(1),(LEFT5,IK,P(IK),IK=2,N1M1)
 3280 FORMAT(17X,'X(',I3,11H) = Z( 1)**,I1,9(A5,I2,3H)**,I1)/(23X,A5,I2,
     13H)**,I1,A5,I2,3H)**,I1,A5,I2,3H)**,I1,A5,I2,3H)**,I1,A5,I2,3H)**,
     2I1,A5,I2,3H)**,I1,A5,I2,3H)**,I1,A5,I2,3H)**,I1,A5,I2,3H)**,I1,A5,
     3I2,3H)**,I1))
      BZERO=XBAR(N)-BZERO
      R2=(1-A(N,N))*100.0
      WRITE(3,3290)BZERO,R2
 3290 FORMAT(3X,'CONSTANT =',F12.5/3X,'MULTIPLE CORRELATION COEFFICIENT
     1SQUARED =',F6.2,' PERCENT')
      FTEST2=ABS(VMIN)*DF/A(N,N)
      IF(FTEST2.GE.F2) GO TO 480
      WRITE(3,3300)STEPNO,MIN,FTEST2
 3300 FORMAT(///1X,'STEP NO.',I3/3X,'VARIABLE REMOVED',I4/
     13X,'F LEVEL',F8.4)
      K=MIN
```

```
      DF=DF+1
      GO TO 560
 480  FTEST1=VMAX*(DF-1)/(A(N,N)-VMAX)
      IF(FTEST1.GT.F1) GO TO 550
      REWIND 4
      IF(IRESID.EQ.0) GO TO 10
      WRITE(3,3310)
3310  FORMAT(3X,'PREDICTED VS. ACTUAL RESULTS'/10X,'OBSERVATION',10X,
     1'ACTUAL',11X,'PREDICTED',11X,'DEVIATION')
      DO 540 IK=1,M
      READ(4,FMT)(Z(J),J=1,N1)
      TEMP=Z(DEPVAR)
      DO 485 IL=DEPVAR,N1M1
 485  Z(IL)=Z(IL+1)
      Z(N1)=TEMP
      SUMPR=0.0
      DO 530 J=1,INEQ
      IF(MODE.EQ.INORM) GO TO 490
      IRET=LIST(INVAR(J))
      GO TO 500
 490  IRET=INVAR(J)
 500  DO 510 IJ=1,N1M1
      IZ=IRET/(MAXP(N1-IJ)+1)
      P(N1-IJ)=IRET-IZ*(MAXP(N1-IJ)+1)
 510  IRET=IZ
      PROD=1.0
      DO 520 II=1,N1M1
      IF(P(II).EQ.0) GO TO 520
      PROD=PROD*Z(II)**P(II)
 520  CONTINUE
 530  SUMPR=SUMPR+PROD*B(J)
```

```
      YHAT=BZERO+SUMPR
      DEV=Z(N1)-YHAT
  540 WRITE(3, 3320)IK, Z(N1), YHAT, DEV
 3320 FORMAT(12X, I5, 4X, F16. 5, 4X, F16. 5, 4X, F16. 5)
      REWIND 4
      GO TO 10
  550 WRITE(3, 3330)STEPNO, MAX, FTEST1
 3330 FORMAT(///1X, 'STEP NO. ', I3/3X, 'VARIABLE ENTERING', I4/
     13X, 'F LEVEL', F8. 4)
      K=MAX
      DF=DF-1
  560 PIVOT=A(K, K)
      DO 590 I=1, N
      IF(I. EQ. K) GO TO 590
      PIVCOL=A(I, K)
      DO 580 J=1, N
      IF(J. EQ. K) GO TO 570
      A(I, J)=(A(I, J)*PIVOT-PIVCOL*A(K, J))/PIVOT
      GO TO 580
  570 A(I, J)=-A(I, K)/PIVOT
  580 CONTINUE
  590 CONTINUE
      DO 610 J=1, N
      IF(J. EQ. K) GO TO 600
      A(K, J)=A(K, J)/PIVOT
      GO TO 610
  600 A(K, K)=1. 0/PIVOT
  610 CONTINUE
      GO TO 380
  620 CALL EXIT
      END
```

```
      SUBROUTINE GENLST(MODE, N1M1, N, NM1, NCARD, MAXP, LIST, ELM, ALPHA)
      DATA IBLK/4H    /, ICOM/4H,   /, IAST/4H*   /, ISHIFT/Z01000000/,
     1IACCE/4HACCE/, IDELE/4HDELE/
      INTEGER P(15), MAXP(15), ALPHA(3000), ELM(200, 15), LIST(150)
      K=NCARD*80
      READ(1, 1000)(ALPHA(J), J=1, K)
 1000 FORMAT(80A1)
      IROW=1
      ICOL=0
      DO 40 I=1, K
      IF(ALPHA(I). EQ. IBLK) GO TO 40
      IF(ALPHA(I). NE. ICOM) GO TO 20
      IF(ICOL. EQ. N1M1) GO TO 10
      WRITE(3, 3000)
 3000 FORMAT(' INCORRECT NUMBER OF CHARACTERS BETWEEN COMMAS ON A VARIAB
     1LE MODE CARD')
      STOP
   10 IROW=IROW+1
      ICOL=0
      GO TO 40
   20 ICOL=ICOL+1
      IF(ALPHA(I). EQ. IAST) GO TO 30
      ELM(IROW, ICOL)=15-IABS(ALPHA(I)/ISHIFT)
      GO TO 40
   30 ELM(IROW, ICOL)=IAST
   40 CONTINUE
      DO 50 I=1, N1M1
   50 P(I)=0
      P(N1M1)=1
      I=0
      LL=0
```

```
 60 ITEST=N1M1
 70 I=I+1
    DO 90 II=1,IROW
    DO 80 IJ=1,ICOL
    IF(ELM(II,IJ).EQ.IAST) GO TO 80
    IF(ELM(II,IJ).NE.P(IJ)) GO TO 90
 80 CONTINUE
    GO TO 100
 90 CONTINUE
    IF(MODE.EQ.IACCE) GO TO 120
    GO TO 110
100 IF(MODE.EQ.IDELE) GO TO 120
110 LL=LL+1
    LIST(LL)=I
120 IF(P(ITEST).EQ.MAXP(ITEST)) GO TO 130
    P(ITEST)=P(ITEST)+1
    GO TO 70
130 IF(P(ITEST-1).NE.MAXP(ITEST-1)) GO TO 140
    IF(ITEST.EQ.2) GO TO 160
    ITEST=ITEST-1
    GO TO 130
140 DO 150 II=ITEST,N1M1
150 P(II)=0
    P(ITEST-1)=P(ITEST-1)+1
    GO TO 60
160 NM1=LL
    N=LL+1
    RETURN
    END
```

```
      SUBROUTINE GENTRM(N1M1, NM1, MAXP, LIST, Z, X)
      DIMENSION Z(15), X(150)
      INTEGER P(15), MAXP(15), LIST(150)
      N1=N1M1+1
      DO 20 I=1, NM1
      IRET=LIST(I)
      DO 10 IJ=1, N1M1
      IZ=IRET/(MAXP(N1-IJ)+1)
      P(N1-IJ)=IRET-IZ*(MAXP(N1-IJ)+1)
   10 IRET=IZ
      X(I)=1.0
      DO 20 II=1, N1M1
      IF(P(II).EQ.0) GO TO 20
      X(I)=X(I)*Z(II)**P(II)
   20 CONTINUE
      X(NM1+1)=Z(N1)
      RETURN
      END
```

```
      SUBROUTINE EXPAND(N1M1, MAXP, Z, X)
      DIMENSION Z(15), X(125)
      INTEGER P(15), MAXP(15)
      DO 10 I=1, N1M1
   10 P(I)=0
      P(N1M1)=1
      I=0
   20 ITEST=N1M1
   30 I=I+1
      X(I)=1.0
      DO 40 II=1, N1M1
      IF(P(II).EQ.0) GO TO 40
      X(I)=X(I)*Z(II)**P(II)
   40 CONTINUE
      IF(P(ITEST).EQ.MAXP(ITEST)) GO TO 50
      P(ITEST)=P(ITEST)+1
      GO TO 30
   50 IF(P(ITEST-1).NE.MAXP(ITEST-1)) GO TO 60
      IF(ITEST.EQ.2) GO TO 80
      ITEST=ITEST-1
      GO TO 50
   60 DO 70 II=ITEST, N1M1
   70 P(II)=0
      P(ITEST-1)=P(ITEST-1)+1
      GO TO 20
   80 X(I+1)=Z(N1M1+1)
      RETURN
      END
```

## X. COMMENTED PROGRAM LISTING

```
C     MULTIPLE REGRESSION ANALYSIS (STEPWISE)
C     A(I,J)=(N BY N) SQUARE MATRIX ON WHICH LINEAR TRANSFORMATIONS ARE
C         PERFORMED IN ORDER TO OBTAIN THE REGRESSION EQ. AT EACH STEP
C     B(I)=ESTIMATED COEFFICIENT OF A INDEPENDENT TERM - NOT NECESSARILY
C         TERM I
C     BZERO=CONSTANT TERM
C     CONTRL=INTEGER VARIABLE WHOSE VALUE IS THE BINARY REPRESENTATION
C         OF ALPHA CHARACTERS USED TO DETECT INCORRECT OR MISSING
C         CONTROL CARDS
C     DEPVAR=VALUE OF THE NUMBER OF THE INPUT VARIABLE THAT IS
C         DESIGNATED AS THE DEPENDENT VARIABLE
C     DEV=DEVIATION BETWEEN THE OBSERVED AND THE PREDICTED VALUE OF THE
C         DEPENDENT VARIABLE
C     DF=DEGREES OF FREEDOM
C     ERRORC(I)=STANDARD ERROR OF COEFFICIENT OF VARIABLE I
C     ERRORY=STANDARD ERROR OF DEPENDENT VARIABLE
C     EXPAND=SUBROUTINE THAT GENERATES THE VALUES OF ALL THE TERMS FOR A
C         GIVEN OBSERVATION WHEN IN NORMAL MODE
C     FMT(I)=ARRAY OF ALPHA INFORMATION THAT INDICATES THE FORMAT TO
C         USE WHEN READING THE INPUT DATA
C     FTEST1=CALCULATED F VALUE USED TO TEST FOR ENTRANCE OF A VARIABLE
C     FTEST2=CALCULATED F VALUE USED TO TEST FOR REMOVAL OF A VARIABLE
C     F1=F VALUE FOR ENTERING VARIABLE
C     F2=F VALUE FOR REMOVING VARIABLE
C     GENLST=SUBROUTINE THAT GENERATES A LIST OF TERM NUMBERS - WHEN IN
C         THE NORMAL MODE MANY COMBINATIONS OF THE POWERS OF THE
C         INDEPENDENT VARIABLES ARE POSSIBLE WITHOUT EXCEEDING THE
C         MAXIMUM POWER OF ANY INDEPENDENT VARIABLE - EACH POSSIBLE
C         COMBINATION REPRESENTS ONE INDEPENDENT TERM - ALL POSSIBLE
```

```
C        COMBINATIONS MAKE UP A COMPLETE SET OF TERMS - THESE
C        INDEPENDENT TERMS ARE GENERATED IN AN ORDERLY FASHION SO THAT
C        EACH INDEPENDENT TERM IS GIVEN A CONSECUTIVE NUMBER - WHEN IN
C        ACCEPT OR DELETE MODE, ONLY A SUBSET OF THE COMPLETE SET OF
C        TERMS IS DESIRED - THUS THE LIST IS COMPOSED OF THE NUMBER OF
C        THE TERMS IN A COMPLETE SET THAT CORRESPOND TO THE DESIRED
C        TERMS
C     GENTRM=SUBROUTINE THAT GENERATES THE VALUES OF ALL THE DESIRED
C        TERMS FOR A GIVEN OBSERVATION WHEN IN DELETE OR ACCEPT MODE
C     I=TERM NUMBER AND COUNTER
C     II=DO LOOP PARAMETER
C     IJ=DO LOOP PARAMETER
C     IK=DO LOOP PARAMETER
C     IL=DO LOOP PARAMETER
C     ILIST=SWITCH THAT INDICATES WHETHER THE LIST THAT TRANSFORMS
C        DELETE OR ACCEPT MODE TO NORMAL MODE SHOULD BE PRINTED - A
C        VALUE OF ZERO INDICATES NO PRINT OUT
C     INDATA=SWITCH THAT INDICATES WHETHER THE INPUT DATA SHOULD BE
C        PRINTED, A VALUE OF ZERO INDICATES NO PRINT-OUT
C     INEQ=SUBSCRIPT OF THE VECTOR THAT CONTAINS THE INDEPENDENT TERMS
C        IN THE REGRESSION EQUATION
C     INPTAP(I)=ARRAY USED TO TRANSFER CARD IMAGES FROM INPUT DATA CARDS
C        TO THE LOGICAL UNIT DESIGNATED
C     INVAR(I)=VECTOR THAT CONTAINS THE VALUE OF THE NUMBER OF THE TERMS
C        PRESENTLY IN THE REGRESSION EQUATION
C     IP1=LOWER BOUND ON A DO LOOP
C     IR=SWITCH THAT INDICATES WHETHER THE CORRELATION COEFFICIENTS
C        SHOULD BE PRINTED, A VALUE OF ZERO INDICATES NO PRINT-OUT
C     IRESID=SWITCH THAT INDICATES WHETHER THE RESIDUALS SHOULD BE
C        PRINTED, A VALUE OF ZERO INDICATES NO PRINT-OUT
C     IRET=NUMBER OF THE TERM IN NORMAL MODE THAT IS TO BE RETRIEVED AS
```

```
C         POWERS OF ITS INDEPENDENT VARIABLES
C     IS=SWITCH THAT INDICATES WHETHER THE WEIGHTED RESIDUAL SUMS OF
C         SQUARES AND CROSS PRODUCTS SHOULD BE PRINTED, A VALUE OF ZERO
C         INDICATES NO PRINT-OUT
C     ISMSCP=SWITCH THAT INDICATES WHETHER THE WEIGHTED SUMS OF SQUARES
C         AND CROSS PRODUCTS SHOULD BE PRINTED, A VALUE OF ZERO
C         INDICATES NO PRINT-OUT
C     ISUMVR=SWITCH THAT INDICATES WHETHER THE WEIGHTED SUMS OF
C         VARIABLES SHOULD BE PRINTED, A VALUE OF ZERO INDICATES NO
C         PRINT-OUT
C     IW=SWITCH THAT INDICATES WHETHER THE WEIGHTS OF ALL OBSERVATIONS
C         ARE CONSTANT OR VARIABLE, A VALUE OF ONE INDICATES ALL
C         OBSERVATIONS HAVE EQUAL WEIGHTS
C     IXBAR=SWITCH THAT INDICATES WHETHER THE WEIGHTED MEAN OF VARIABLES
C         SHOULD BE PRINTED, A VALUE OF ZERO INDICATES NO PRINT-OUT
C     IZ=FRACTIONAL PART OF IRET TO STILL BE RETRIEVED
C     J=DO LOOP PARAMETER
C     K=ROW AND COLUMN OF MATRIX A THAT CORRESPONDS TO THE INDEPENDENT
C         TERM THAT IS TO BE ADDED OR REMOVED
C     LEFT1=INTEGER VARIABLE WHOSE VALUE IS THE BINARY REPRESENTATION OF
C         THE TWO CHARACTERS X-
C     LEFT2=INTEGER VARIABLE WHOSE VALUE IS THE BINARY REPRESENTATION OF
C         THE TWO CHARACTERS Z(
C     LEFT3=INTEGER VARIABLE WHOSE VALUE IS THE BINARY REPRESENTATION OF
C         THE SIX CHARACTERS SUM X(
C     LEFT4=INTEGER VARIABLE WHOSE VALUE IS THE BINARY REPRESENTATION OF
C         THE TWO CHARACTERS X(
C     LEFT5=INTEGER VARIABLE WHOSE VALUE IS THE BINARY REPRESENTATION OF
C         THE FIVE CHARACTERS  * Z(
C     LIST(I)=VALUE OF THE NUMBER OF THE INDEPENDENT TERM GENERATED IN
C         THE NORMAL MODE THAT CORRESPONDS TO THE NUMBER OF THE
```

```
C          INDEPENDENT TERM GENERATED IN THE MODE OF THE PROGRAM BEING
C          PROCESSED
C      M=NUMBER OF OBSERVATIONS
C      MAX=THE NUMBER OF THE INDEPENDENT TERM THAT WOULD CAUSE THE
C          GREATEST VARIANCE REDUCTION WHEN PUT INTO REGRESSION
C      MAXP(I)=MAXIMUM POWER OF INDEPENDENT VARIABLE I ALLOWED IN ANY
C          INDEPENDENT TERM
C      MEANSS=SUM OF SQUARES OF THE MEAN IN THE ANALYSIS OF VARIANCE
C          TABLE
C      MIN=THE NUMBER OF THE INDEPENDENT TERM THAT WOULD CAUSE THE LEAST
C          VARIANCE INCREASE WHEN TAKEN OUT OF REGRESSION
C      MODE=ALPHA INFORMATION USED TO DETECT THE VARIABLE MODE
C      N=NUMBER OF INDEPENDENT TERMS + THE ONE DEPENDENT TERM
C      NCARD=NUMBER OF VARIABLE MODE CONTROL DATA CARDS
C      NM1=NUMBER OF INDEPENDENT TERMS
C      NODATA=WEIGHTED NUMBER OF DATA
C      N1=NUMBER OF INDEPENDENT VARIABLES  + ONE FOR THE DEPENDENT
C          VARIABLE
C      N1M1=NUMBER OF INDEPENDENT VARIABLES
C      P(I)=POWER OF THE INDEPENDENT VARIABLE I NEEDED TO COMPOSE THE
C          INDEPENDENT TERM BEING CONSIDERED
C      PIVCOL=PIVOT COLUMN OF MATRIX A
C      PIVOT=PIVOT ELEMENT OF MATRIX A
C      PROD=CUMULATIVE PRODUCT OF INDEPENDENT VARIABLES TIMES THEIR
C          RESPECTIVE POWERS FOR THE INDEPENDENT TERM BEING CALCULATED
C      PRONUM=PROBLEM NUMBER
C      R(I, J)=SIMPLE CORRELATION COEFFICIENT OF VARIABLES I AND J
C      RATIO=F RATIO IN THE ANALYSIS OF VARIANCE TABLE
C      REGDF=DEGREES OF FREEDOM ATTRIBUTED TO REGRESSION
C      REGMS=REGRESSION MEAN SQUARE IN THE ANALYSIS OF VARIANCE TABLE
C      REGSS=REGRESSION SUM OF SQUARES IN THE ANALYSIS OF VARIANCE TABLE
```

```
C     RESMS=RESIDUAL MEAN SQUARE IN THE ANALYSIS OF VARIANCE TABLE
C     RESSS=RESIDUAL SUM OF SQUARES IN THE ANALYSIS OF VARIANCE TABLE
C     R2=MULTIPLE CORRELATION COEFFICIENT SQUARED
C     S(I, J)=WEIGHTED RESIDUAL SUM OF SQUARES AND CROSS PRODUCTS OF
C         TERMS I AND J
C     SIGMA(I)=SQUARE ROOT OF S(I, I)
C     STEPNO=STEP NUMBER - NUMBER OF DIFFERENT REGRESSION EQUATIONS
C         CALCULATED
C     SUMPR=CUMULATIVE SUM OF EACH INDEPENDENT TERM TIMES ITS
C         CORRESPONDING REGRESSION COEFFICIENT FOR A GIVEN OBSERVATION
C     SUMSCP(I, J)=WEIGHTED SUM OF SQUARES AND CROSS PRODUCTS OF TERMS I
C         AND J
C     SUMVAR(I)=WEIGHTED SUMS OF TERM I
C     TEMP=TEMPORARY STORAGE AREA
C     TOALSS=TOTAL SUM OF SQUARES IN THE ANALYSIS OF VARIANCE TABLE
C     TOL=TOLERANCE TO CONTROL THE POSSIBILITY OF DEGENERACY
C     V=ADDITIONAL VARIANCE ATTRIBUTED TO THE INDEPENDENT TERM BEING
C         CONSIDERED
C     VMAX=VARIANCE REDUCTION BY ADDING A TERM - TERM NUMBER EQUALS MAX
C     VMIN=VARIANCE INCREASE BY DELETING A TERM - TERM NUMBER EQUALS MIN
C     WEIGHT=WEIGHTING FACTOR FOR A PARTICULAR OBSERVATION
C     X(I)=INDEPENDENT TERM I FOR A SPECIFIED OBSERVATION - IT IS
C         ONE OF MANY COMBINATIONS OF THE INDEPENDENT VARIABLES
C     X(N)=Z(N1)
C     XBAR(I)=WEIGHTED MEAN OF TERM I
C     YHAT=PREDICTED VALUE OF THE DEPENDENT VARIABLE
C     Z(I)=INDEPENDENT VARIABLE I, (I=1, N1M1), FOR A SPECIFIC
C         OBSERVATION
C     Z(N1)=THE DEPENDENT VARIABLE FOR A SPECIFIC OBSERVATION
C
      REAL Z(15), X(150), SUMVAR(150), XBAR(150), SUMSCP(150, 150), S(150, 150)
```

```
     1,SIGMA(150),R(150,150),A(150,150),NODATA,MEANSS,B(150)
      INTEGER DF,STEPNO,PRONUM,MAXP(15),P(15),REGDF,LIST(150),CONTRL,
     1DEPVAR,FMT(18),INVAR(150),INPTAP(20)
      DOUBLE PRECISION LEFT3,LEFT5
      EQUIVALENCE (SUMVAR,XBAR),(SUMSCP,S,R,A),(X,SIGMA)
      DATA LEFT1/2HX-/,LEFT2/2HZ(/,LEFT3/6HSUM X(/,LEFT4/2HX(/,
     1LEFT5/5H * Z(/,IFORM/4HFORM/,ISTAR/4HSTAR/,IEND/4HEND /,
     2INFO/4HINFO/,IPRIN/4HPRIN/,IMAX/4HMAX /,INORM/4HNORM/,
     3IACCE/4HACCE/,IDELE/4HDELE/
C
C     SET THE TOLERANCE TO CONTROL THE POSSIBILITY OF DEGENERACY
C
      TOL=.001
C
C     READ CONTROL CARD - EITHER START OF INPUT DATA CARD OR INFORMATION
C         CARD
C
   10 READ(1,1000,END=620)CONTRL,PRONUM,N1,F1,F2,IW
 1000 FORMAT(A4,6X,2I5,2F5.2,I5)
C
C     IF THE LAST CARD READ WAS A START OF INPUT DATA CONTROL CARD, THEN
C         WRITE THE CARD IMAGES OF THE FOLLOWING INPUT DATA CARDS ON A
C         SPECIFIED LOGICAL UNIT (TAPE OR DISK) - OTHERWISE MAKE
C         FURTHER TESTS ON THE CARD IN QUESTION
C
   20 IF(CONTRL.NE.ISTAR) GO TO 50
C
C     READ A DATA CARD AND PLACE THE CARD IMAGE AS ALPHA INFORMATION IN
C         ARRAY INPTAP
C
   30 READ(1,1010)INPTAP
```

```
 1010 FORMAT(20A4)
C
C     IF THE MOST RECENTLY READ CARD WAS A DATA CARD, THEN WRITE THE
C          CARD IMAGE CONTAINED IN ARRAY INPTAP ON THE SPECIFIED UNIT -
C          OTHERWISE THE MOST RECENT CARD WAS AN END OF INPUT DATA
C          CONTROL CARD AND AN END OF FILE MARK IS WRITTEN ON THE UNIT
C          SPECIFIED
C
      IF(INPTAP(1). EQ. IEND) GO TO 40
C
C     WRITE THE CARD IMAGE ON THE LOGICAL UNIT SPECIFIED
C
      WRITE(4, 1010)INPTAP
C
C     GO READ THE NEXT CARD - IT WILL BE A DATA CARD OR AN END OF INPUT
C          DATA CONTROL CARD
C
      GO TO 30
C
C     WRITE AN END OF FILE MARK ON THE UNIT SPECIFIED
C
   40 ENDFILE 4
      REWIND 4
C
C     GO READ THE NEXT CONTROL CARD - INFORMATION CARD
C
      GO TO 10
C
C     CHECK WHETHER THE INFORMATION CONTROL CARD IS CORRECT
C
   50 IF(CONTRL. EQ. INFO) GO TO 60
```

```
C
C     PRINT MESSAGE THAT THE INFORMATION CONTROL CARD IS INCORRECT
C
      WRITE(3,3000)
 3000 FORMAT(' INCORRECT OR MISSING CONTROL CARD - EITHER INFORMATION CA
     1RD OR START OF INPUT DATA CARD')
      GO TO 620
C
C     PRINT INFORMATION FROM THE INFORMATION CONTROL CARD
C
   60 WRITE(3,3010)PRONUM,N1,F1,F2
 3010 FORMAT('1PROBLEM NO',I5//' NO OF VARIABLES =',I5//' F LEVEL TO ENT
     1ER VARIABLE =',F5.2//' F LEVEL TO REMOVE VARIABLE =',F5.2)
C
C     READ PRINT CONTROL CARD
C
      READ(1,1020)CONTRL,INDATA,ISUMVR,IXBAR,ISMSCP,IS,IR,ILIST,IRESID
 1020 FORMAT(A4,6X,14I5)
C
C     CHECK WHETHER THE PRINT CONTROL CARD IS CORRECT
C
      IF(CONTRL.EQ.IPRIN) GO TO 70
C
C     PRINT MESSAGE THAT PRINT CONTROL CARD IS INCORRECT
C
      WRITE(3,3020)
 3020 FORMAT(' INCORRECT OR MISSING CONTROL CARD - PRINT CARD')
      GO TO 620
C
C     COMPUTE THE NUMBER OF INDEPENDENT VARIABLES (N1M1)
C
```

```
   70 N1M1=N1-1
C
C     READ MAXIMUM POWERS CONTROL CARD
C
      READ(1, 1020)CONTRL, (MAXP(I), I=1, N1M1)
C
C     CHECK WHETHER THE MAXIMUM POWERS CONTROL CARD IS CORRECT
C
      IF(CONTRL. EQ. IMAX) GO TO 80
C
C     PRINT MESSAGE THAT THE MAXIMUM POWERS CONTROL CARD IS INCORRECT
C
      WRITE(3, 3030)
 3030 FORMAT(' INCORRECT OR MISSING CONTROL CARD - MAXIMUM POWERS CARD')
      GO TO 620
C
C     READ FORMAT CONTROL CARD
C
   80 READ(1, 1030)CONTRL, FMT, DEPVAR
 1030 FORMAT(A4, 2X, 18A4, I2)
C
C     CHECK WHETHER THE FORMAT CONTROL CARD IS CORRECT
C
      IF(CONTRL. EQ. IFORM) GO TO 90
C
C     PRINT MESSAGE THAT FORMAT CONTROL CARD IS INCORRECT
C
      WRITE(3, 3040)
 3040 FORMAT(' INCORRECT OR MISSING CONTROL CARD - FORMAT CARD')
      GO TO 620
C
```

```
C     READ THE VARIABLE MODE CONTROL CARD
C
   90 READ(1, 1020)MODE, NCARD
C
C     DETERMINE THE VARIABLE MODE (ACCEPT, DELETE, OR NORMAL)
C
      IF(MODE. EQ. INORM) GO TO 110
      IF((MODE. EQ. IDELE). OR. (MODE. EQ. IACCE)) GO TO 100
C
C     PRINT MESSAGE THAT THE VARIABLE MODE CONTROL CARD IS INCORRECT
C
      WRITE(3, 3050)
 3050 FORMAT(' INCORRECT OR MISSING CONTROL CARD - VARIABLE MODE CARD')
      GO TO 620
C
C     GENERATE A LIST OF INDEPENDENT TERM NUMBERS FOR THE VARIABLE MODE
C          BEING CONSIDERED - THE LIST IS COMPOSED OF A SUBSET OF THE
C          COMPLETE SET OF TERM NUMBERS, AND CONTAINS THE TERM NUMBERS
C          OF THE INDEPENDENT TERMS BEING USED AS IF THEY HAD BEEN
C          GENERATED IN THE NORMAL MODE
C
  100 CALL GENLST(MODE, N1M1, N, NM1, NCARD, MAXP(1), LIST(1), A(1, 1), A(1, 76))
      GO TO 130
C
C     COMPUTE THE NUMBER OF TERMS IN THE NORMAL MODE (N)
C
  110 N=1
      DO 120 I=1, N1M1
  120 N=N*(MAXP(I)+1)
C
C     COMPUTE THE NUMBER OF INDEPENDENT TERMS IN THE NORMAL MODE (NM1)
```

```
C
      NM1=N-1
C
C     PRINT INFORMATION ON THE MAXIMUM POWERS CONTROL CARD
C
  130 WRITE(3,3060)N1M1,NM1,(LEFT1,I,MAXP(I),I=1,N1M1)
 3060 FORMAT(/' NO. OF INDEPENDENT VARIABLES =',I3//' NO. OF INDEPENDENT
     1 TERMS =',I4//3X,'INDEPENDENT',4X,'MAX'/5X,'VARIABLE',4X,'POWER'/
     2(7X,A2,I2,7X,I2))
C
C     CHECK WHETHER THE INPUT DATA SHOULD BE PRINTED
C
      IF(INDATA.EQ.0)GO TO 140
C
C     PRINT HEADING FOR INPUT DATA
C
      WRITE(3,3070)(LEFT2,(I),I=1,N1)
 3070 FORMAT(///20X,'INPUT DATA'//' OBSERVATION',10(3X,A2,I2,')',2X)/
     1(15X,A2,I2,')',5X,A2,I2,')',5X,A2,I2,')',5X,A2,I2,')',5X,A2,I2,')'
     2,5X,A2,I2,')',5X,A2,I2,')',5X,A2,I2,')',5X,A2,I2,')',5X,A2,I2,')')
     3)
C
C     INITIALIZE THE ARRAYS SUMVAR AND SUMSCP TO THE VALUE ZERO
C
  140 DO 150 I=1,N
      SUMVAR(I)=0.0
      DO 150 J=1,N
  150 SUMSCP(I,J)=0.0
C
C     INITIALIZE THE WEIGHTED NUMBER OF DATA, THE WEIGHT, AND THE
C          NUMBER OF OBSERVATIONS
```

```
C
      NODATA=0.0
      WEIGHT=1.0
      M=0
C
C     DETERMINE WHETHER ALL OBSERVATIONS HAVE THE SAME WEIGHT
C
  160 IF(IW.EQ.0) GO TO 162
C
C     READ ONE OBSERVATION OF THE INPUT DATA FROM THE SPECIFIED LOGICAL
C         UNIT WITH THE FORMAT OF THE FORMAT CONTROL CARD
C
      READ(4,FMT,END=240)(Z(J),J=1,N1)
      GO TO 170
C
C     READ ONE OBSERVATION OF THE INPUT DATA AND THE WEIGHT OF THAT
C         OBSERVATION FROM THE SPECIFIED LOGICAL UNIT WITH THE FORMAT
C         OF THE FORMAT CONTROL CARD
C
  162 READ(4,FMT,END=240)(Z(J),J=1,N1),WEIGHT
C
C     INCREMENT THE NUMBER OF OBSERVATIONS BY ONE
C
  170 M=M+1
C
C     MOVE THE VALUE OF THE DEPENDENT VARIABLE TO A TEMPORARY STORAGE
C         AREA
C
      TEMP=Z(DEPVAR)
C
C     MOVE THE VALUE OF EACH INPUT VARIABLE POSITIONED BEYOND THE
```

```
C          DEPENDENT VARIABLE TO THE LOCATION OF THE INPUT VARIABLE
C          THAT PRECEDED IT
C
      DO 180 J=DEPVAR, N1M1
  180 Z(J)=Z(J+1)
C
C     MOVE THE VALUE OF THE DEPENDENT VARIABLE TO THE LOCATION OF THE
C          LAST INPUT VARIABLE
C
      Z(N1)=TEMP
C
C     CHECK WHETHER THE INPUT DATA OF THE OBSERVATION BEING CONSIDERED
C          SHOULD BE PRINTED
C
      IF(INDATA. EQ. 0)GO TO 190
C
C     PRINT THE INPUT DATA OF THE OBSERVATION BEING CONSIDERED
C
      WRITE(3, 3080)M, (Z(J), J=1, N1)
 3080 FORMAT(3X,I4, 4X, 10F10. 3, (/11X, 10F10. 3))
C
C     INCREMENT THE WEIGHTED NUMBER OF DATA BY THE WEIGHT
C
  190 NODATA=NODATA+WEIGHT
C
C     IF THE MODE OF THE PROGRAM IS NORMAL, THEN CALL SUBROUTINE EXPAND
C          TO GENERATE A COMPLETE SET OF TERMS - OTHERWISE CALL
C          SUBROUTINE GENTRM TO GENERATE A SUBSET OF THE COMPLETE SET
C          OF TERMS
C
      IF(MODE. EQ. INORM) GO TO 200
```

```
C
C     GENERATE THE TERMS FOR THE OBSERVATION AND MODE (ACCEPT OR DELETE)
C         BEING CONSIDERED
C
      CALL GENTRM(N1M1, NM1, MAXP(1), LIST(1), Z(1), X(1))
      GO TO 210
C
C     GENERATE ALL TERMS IN NORMAL MODE FOR THE OBSERVATION BEING
C         CONSIDERED
C
  200 CALL EXPAND(N1M1, MAXP(1), Z(1), X(1))
  210 DO 220 J=1, N
C
C     UPDATE THE CUMULATIVE WEIGHTED SUM OF TERMS BY ADDING THE
C         CONTRIBUTION OF THE OBSERVATION BEING CONSIDERED
C
  220 SUMVAR(J)=SUMVAR(J)+WEIGHT*X(J)
      DO 230 I=1, N
      DO 230 J=I, N
C
C     UPDATE THE CUMULATIVE WEIGHTED SUM OF SQUARES AND CROSS PRODUCTS
C         BY ADDING THE CONTRIBUTION OF THE OBSERVATION BEING
C         CONSIDERED
C
  230 SUMSCP(I, J)=SUMSCP(I, J)+WEIGHT*X(I)*X(J)
      GO TO 160
C
C     PRINT THE NUMBER OF OBSERVATIONS AND THE WEIGHTED NUMBER OF DATA
C
  240 WRITE(3, 3090)M, NODATA
 3090 FORMAT(/' NO. OF OBSERVATIONS =',I5//' WEIGHTED NO. OF DATA =',
```

```
     1F7.2)
C
C     CHECK WHETHER THE WEIGHTED SUM OF TERMS SHOULD BE PRINTED
C
      IF(ISUMVR.EQ.0)GO TO 250
C
C     PRINT THE WEIGHTED SUM OF TERMS
C
      WRITE(3,3100)
 3100 FORMAT(///20X,'WEIGHTED SUMS OF VARIABLES'/)
      WRITE(3,3110)((LEFT3,I,SUMVAR(I)),I=1,NM1)
 3110 FORMAT(3(1X,A6,I2,') =',F14.5,4X))
      WRITE(3,3120)SUMVAR(N)
 3120 FORMAT(' SUM    Y    =',F14.5)
  250 DO 260 I=1,N
C
C     COMPUTE THE WEIGHTED MEAN OF ALL THE TERMS
C
  260 XBAR(I)=SUMVAR(I)/NODATA
C
C     CHECK WHETHER THE WEIGHTED MEANS SHOULD BE PRINTED
C
      IF(IXBAR.EQ.0)GO TO 270
C
C     PRINT THE WEIGHTED MEANS
C
      WRITE(3,3130)
 3130 FORMAT(///20X,'WEIGHTED MEAN OF VARIABLES'/)
      WRITE(3,3110)((LEFT3,I,XBAR(I)),I=1,NM1)
      WRITE(3,3120)XBAR(N)
C
```

```
C     CHECK WHETHER THE WEIGHTED SUM OF SQUARES AND CROSS PRODUCTS
C          SHOULD BE PRINTED
C
  270 IF(ISMSCP. EQ. 0) GO TO 280
C
C     PRINT THE WEIGHTED SUM OF SQUARES AND CROSS PRODUCTS
C
      WRITE(3, 3140)
 3140 FORMAT(///20X, 'WEIGHTED SUMS OF SQUARES AND CROSS PRODUCTS'/)
      WRITE(3, 3150)(((LEFT4, I, J, SUMSCP(I, J)), J=I, NM1), I=1, NM1)
 3150 FORMAT(3(1X, A2, I2, ') VS X(', I2, ') =', F16. 5, 4X))
      WRITE(3, 3160)((LEFT4, I, SUMSCP(I, N)), I=1, NM1)
 3160 FORMAT(3(1X, A2, I2, ') VS     Y    =', F16. 5, 4X))
      WRITE(3, 3170)SUMSCP(N, N)
 3170 FORMAT(' Y       VS    Y    =', F16. 5)
C
C     COMPUTE THE TOTAL SUM OF SQUARES OF THE DEPENDENT VARIABLE FOR
C          THE ANALYSIS OF VARIANCE TABLE
C
  280 TOALSS=SUMSCP(N, N)
      DO 300 I=1, N
      DO 290 J=I, N
C
C     COMPUTE THE WEIGHTED RESIDUAL SUM OF SQUARES AND CROSS PRODUCTS
C          FOR ALL TERMS
C
  290 S(I, J)=SUMSCP(I, J)-XBAR(I)*XBAR(J)*NODATA
  300 SIGMA(I)=SQRT(S(I, I))
C
C     CHECK WHETHER THE WEIGHTED RESIDUAL SUM OF SQUARES AND CROSS
C          PRODUCTS SHOULD BE PRINTED
```

```
C
      IF(IS.EQ.0) GO TO 310
C
C     PRINT THE WEIGHTED RESIDUAL SUM OF SQUARES AND CROSS PRODUCTS
C
      WRITE(3,3180)
 3180 FORMAT(///20X,'WEIGHTED RESIDUAL SUM OF SQUARES AND CROSS PRODUCTS
     1'/)
      WRITE(3,3150)(((LEFT4,I,J,S(I,J)),J=1,NM1),I=1,NM1)
      WRITE(3,3160)((LEFT4,I,S(I,N)),I=1,NM1)
      WRITE(3,3170)S(N,N)
C
C     COMPUTE THE SUM OF SQUARES OF THE MEAN FOR THE AOV TABLE
C
  310 MEANSS=TOALSS-S(N,N)
      DO 320 I=1,NM1
      IP1=I+1
      DO 320 J=IP1,N
C
C     COMPUTE THE STRICTLY UPPER TRIANGULAR PART OF THE MATRIX OF
C         SIMPLE CORRELATION COEFFICIENTS
C
  320 R(I,J)=S(I,J)/(SIGMA(I)*SIGMA(J))
      DO 330 I=1,NM1
      IP1=I+1
      DO 330 J=IP1,N
C
C     COMPUTE THE STRICTLY LOWER TRIANGULAR PART OF THE MATRIX OF SIMPLE
C         CORRELATION COEFFICIENTS
C
  330 R(J,I)=R(I,J)
```

```
      DO 340 I=1,N
C
C     COMPUTE THE DIAGONAL ELEMENTS OF THE MATRIX OF SIMPLE CORRELATION
C         COEFFICIENTS
C
  340 R(I,I)=1.0
C
C     CHECK WHETHER THE MATRIX OF SIMPLE CORRELATION COEFFICIENTS SHOULD
C         BE PRINTED
C
      IF(IR.EQ.0) GO TO 350
C
C     PRINT THE MATRIX OF SIMPLE CORRELATION COEFFICIENTS
C
      WRITE(3,3190)
 3190 FORMAT(///20X,'CORRELATION COEFFICIENTS'/)
      WRITE(3,3150)(((LEFT4,I,J,R(I,J)),J=I,NM1),I=1,NM1)
      WRITE(3,3160)((LEFT4,I,R(I,N)),I=1,NM1)
C
C     COMPUTE THE DEGREES OF FREEDOM
C
  350 DF=NODATA-1.0
C
C     COMPUTE THE STANDARD ERROR OF Y (THE DEPENDENT VARIABLE)
C
      ERRORY=SIGMA(N)*SQRT(R(N,N)/DF)
C
C     PRINT THE STANDARD ERROR OF Y
C
      WRITE(3,3200)ERRORY
 3200 FORMAT(/' STANDARD ERROR OF Y =',F8.3)
```

```
C
C     CHECK WHETHER THE LIST OF TERM NUMBERS THAT IS A SUBSET OF THE
C          COMPLETE SET OF TERM NUMBERS SHOULD BE PRINTED - ONLY
C          APPLICIABLE WHEN IN DELETE OR ACCEPT MODE
C
      IF(ILIST.EQ.0) GO TO 360
C
C     PRINT THE LIST OF CORRESPONDING TERM NUMBERS
C
      WRITE(3,3210)(I,LIST(I),I=1,NM1)
 3210 FORMAT(/' LIST GENERATION'/3X,'(TERM GENERATED) CORRESPONDS TO THE
     1 X TERM GENERATED IN THE NORMAL MODE'//5X,'TERM NO.',3X,'TERM GENE
     2RATED'/(6X,I5,6X,I5))
C
C     INITIALIZE VMAX (VARIANCE REDUCTION) TO ZERO
C
  360 VMAX=0.0
      DO 370 I=1,NM1
C
C     FIND THE SIMPLE CORRELATION COEFFICIENT WITH THE LARGEST ABSOLUTE
C          VALUE - STORE THE VALUE IN VMAX AND THE SUBSCRIPT OF THE
C          CORRESPONDING TERM IN MAX - THIS TERM IS THE MOST HIGHLY
C          CORRELATED WITH THE DEPENDENT VARIABLE (Y)
C
      IF(ABS(R(I,N)).LE.VMAX) GO TO 370
      MAX=I
      VMAX=ABS(R(I,N))
  370 CONTINUE
C
C     COMPUTE THE VARIANCE REDUCTION OF THE MOST HIGHLY CORRELATED TERM
C
```

```
      VMAX=A(MAX, N)*A(N, MAX)
C
C     INITIALIZE THE STEP NUMBER TO THE FIRST STEP
C
      STEPNO=1
      GO TO 480
C
C     INCREMENT THE STEP NUMBER
C
  380 STEPNO=STEPNO+1
C
C     COMPUTE AND PRINT THE STANDARD ERROR OF Y FOR THE MOST RECENT
C         REGRESSION EQUATION
C
      ERRORY=SIGMA(N)*SQRT(A(N, N)/DF)
      WRITE(3, 3220)ERRORY
 3220 FORMAT(3X, 'STANDARD ERROR OF Y =', F12. 4/3X, 'ANALYSIS OF VARIANCE'
     1/12X, 'SOURCE', 6X, 'DF', 4X, 'SUM OF SQUARES', 4X, 'MEAN SQUARE', 3X,
     2'F RATIO')
C
C     PRINT THE TOTAL DEGREES OF FREEDOM AND THE TOTAL SUM OF SQUARES
C         FOR THE ANALYSIS OF VARIANCE TABLE
C
      WRITE(3, 3230)M, TOALSS
 3230 FORMAT(10X, 'TOTAL', 7X, I4, 4X, F14. 3)
C
C     PRINT THE DEGREES OF FREEDOM (1) AND THE SUM OF SQUARES FOR THE
C         MEAN
C
      WRITE(3, 3240)MEANSS
 3240 FORMAT(10X, 'MEAN', 11X, '1', 4X, F14. 3)
```

```
C
C     COMPUTE AND PRINT DEGREES OF FREEDOM (REGDF, RESDF), SUM OF SQUARES
C         (REGSS, RESSS), AND MEAN SQUARES (REGMS, RESMS) FOR REGRESSION
C         AND RESIDUALS - AND COMPUTE AND PRINT THE F RATIO
C
      REGDF=NODATA-DF-1
      RESSS=A(N, N)*(TOALSS-MEANSS)
      REGSS=TOALSS-MEANSS-RESSS
      REGMS=REGSS/REGDF
      RESMS=RESSS/DF
      RATIO=REGMS/RESMS
      WRITE(3, 3250)REGDF, REGSS, REGMS, RATIO
 3250 FORMAT(10X, 'REGRESSION', 2X, I4, 4X, F14. 3, 2X, F13. 3, 2X, F8. 3)
      WRITE(3, 3260)DF, RESSS, RESMS
 3260 FORMAT(10X, 'RESIDUAL', 4X, I4, 4X, F14. 3, 2X, F13. 3/3X, 'VARIABLES IN EQU
     1ATION'/10X, 'VARIABLE', 5X, 'COEFFICIENT', 5X, 'STD ERROR OF COEF')
C
C     INITIALIZE THE NUMBER OF THE TERM BEING CONSIDERED (I), THE
C         SUBSCRIPT FOR THE VECTOR CONTAINING TERMS IN THE REGRESSION
C         EQUATION (INEQ), VARIANCE INCREASE BY DELETING A TERM (VMIN),
C         VARIANCE REDUCTION BY ADDING A TERM (VMAX), NUMBER OF THE
C         TERM CORRESPONDING TO VMIN (MIN), NUMBER OF THE TERM
C         CORRESPONDING TO VMAX (MAX), AND THE CONSTANT TERM (BZERO)
C         EACH TIME A NEW REGRESSION EQUATION IS REVIEWED
C
      I=1
      INEQ=0
      VMIN=5. E10
      VMAX=0. 0
      MIN=0
      MAX=0
```

```
      BZERO=0.0
C
C     TEST WHETHER THE TERM CORRESPONDING TO THE DIAGONAL ELEMENT BEING
C         CONSIDERED IN MATRIX A IS DEGENERATIVE
C
  390 IF(A(I,I).LE.TOL)GO TO 420
C
C     COMPUTE THE ADDITIONAL VARIANCE ATTRIBUTED TO THE TERM BEING
C         CONSIDERED
C
      V=A(I,N)*A(N,I)/A(I,I)
C
C     IF THE TERM BEING CONSIDERED HAS A NEGATIVE VARIANCE CONTRIBUTION
C         THEN PLACE IT IN THE REGRESSION EQUATION
C
      IF(V)410,420,400
C
C     IF THE VARIANCE CONTRIBUTION IS POSITIVE AND GREATER THAN VMAX,
C         THEN ENTER THE VARIANCE CONTRIBUTION INTO VMAX AND ITS
C         CORRESPONDING TERM NUMBER INTO MAX - THIS TERM WILL CAUSE THE
C         GREATEST VARIANCE REDUCTION WHEN INTRODUCED INTO REGRESSION
C
  400 IF(V.LE.VMAX)GO TO 420
      VMAX=V
      MAX=I
      GO TO 420
C
C     INCREMENT THE SUBSCRIPT OF THE VECTOR CONTAINING TERMS IN THE
C         REGRESSION EQUATION - THIS IS ALSO THE NUMBER OF TERMS
C         PRESENTLY IN REGRESSION
C
```

```
  410 INEQ=INEQ+1
C
C     COMPUTE THE COEFFICIENT FOR THIS PARTICULAR TERM IN THE REGRESSION
C         EQUATION
C
      B(INEQ)=A(I, N)*SIGMA(N)/SIGMA(I)
C
C     UPDATE THE CONSTANT TERM BY ADDING THE CONTRIBUTION OF THE TERM
C         BEING CONSIDERED
C
      BZERO=BZERO+B(INEQ)*XBAR(I)
C
C     COMPUTE THE STANDARD ERROR OF THE COEFFICIENT FOR THE TERM BEING
C         CONSIDERED
C
      ERRORC=ERRORY/SIGMA(I)*SQRT(A(I, I))
C
C     PRINT THE NUMBER OF THE TERM BEING CONSIDERED, ITS COEFFICIENT,
C         AND ITS STANDARD ERROR
C
      WRITE(3, 3270)I, B(INEQ), ERRORC
 3270 FORMAT(12X, 'X-', I3, 6X, F10. 5, 7X, F10. 5)
C
C     MOVE THE NUMBER OF THIS TERM THAT IS IN THE REGRESSION EQUATION
C         INTO THE VECTOR INVAR
C
      INVAR(INEQ)=I
C
C     IF THE MAGNITUDE OF THE VARIANCE CONTRIBUTION IS LESS THAN THE
C         MAGNITUDE OF VMIN, THEN ENTER THE VARIANCE CONTRIBUTION INTO
C         VMIN AND ITS CORRESPONDING TERM NUMBER INTO MIN - THIS TERM
```

```
C          WILL CAUSE THE LEAST VARIANCE INCREASE WHEN DELETED FROM
C          REGRESSION
C
      IF(ABS(V).GE.ABS(VMIN))GO TO 420
      VMIN=V
      MIN=I
C
C     IF ALL INDEPENDENT TERMS HAVE BEEN INVESTIGATED, THEN GO CALCULATE
C          THE POWERS OF THE INDEPENDENT VARIABLES FOR THE INDEPENDENT
C          TERMS IN REGRESSION - OTHERWISE INCREMENT THE TERM COUNTER BY
C          ONE AND PROCESS THE NEXT INDEPENDENT TERM
C
  420 IF(I.EQ.NM1) GO TO 430
      I=I+1
      GO TO 390
  430 DO 470 J=1,INEQ
C
C     IF THE NORMAL MODE IS USED, SET IRET EQUAL TO THE TERM NUMBER
C          BEING CONSIDERED (INVAR(J)) - OTHERWISE, SET IRET EQUAL TO
C          THE NUMBER THE TERM BEING CONSIDERED WOULD HAVE HAD IF IT
C          WAS IN THE NORMAL MODE
C
      IF(MODE.EQ.INORM) GO TO 440
      IRET=LIST(INVAR(J))
      GO TO 450
  440 IRET=INVAR(J)
C
C     DETERMINE THE POWERS OF THE INDEPENDENT VARIABLES THAT CORRESPOND
C          TO THE TERM BEING CONSIDERED - THIS IS DONE BY REPEATED
C          DIVISIONS
C
```

```
  450 DO 460 IJ=1,N1M1
C
C     DIVIDE BY (THE MAXIMUM POWER OF THE PARTICULAR INDEPENDENT
C         VARIABLE + ONE) AND THEN TRUNCATE
C
      IZ=IRET/(MAXP(N1-IJ)+1)
C
C     SUBTRACT THE TRUNCATED VALUE FROM THE ORIGINAL VALUE TO OBTAIN
C         THE PARTICULAR POWER
C
      P(N1-IJ)=IRET-IZ*(MAXP(N1-IJ)+1)
C
C     DROP THE PART OF THE TERM ALREADY DIVIDED OUT
C
  460 IRET=IZ
C
C     PRINT THE TERM BEING CONSIDERED AND ITS CORRESPONDING POWERS OF
C         THE INDEPENDENT VARIABLES
C
  470 WRITE(3,3280)INVAR(J), P(1), (LEFT5, IK, P(IK), IK=2, N1M1)
 3280 FORMAT(17X, 'X(', I3, 11H) = Z( 1)**, I1, 9(A5, I2, 3H)**, I1)/(23X, A5, I2,
     13H)**, I1, A5, I2, 3H)**, I1, A5, I2, 3H)**, I1, A5, I2, 3H)**, I1, A5, I2, 3H)**,
     2I1, A5, I2, 3H)**, I1, A5, I2, 3H)**, I1, A5, I2, 3H)**, I1, A5, I2, 3H)**, I1, A5,
     3I2, 3H)**, I1))
C
C     CALCULATE THE CONSTANT TERM
C
      BZERO=XBAR(N)-BZERO
C
C     COMPUTE THE MULTIPLE CORRELATION COEFFICIENT SQUARED
C
```

```
      R2=(1-A(N,N))*100.0
C
C     PRINT THE CONSTANT TERM AND THE MULTIPLE CORRELATION COEFFICIENT
C         SQUARED
C
      WRITE(3,3290)BZERO,R2
 3290 FORMAT(3X,'CONSTANT =',F12.5/3X,'MULTIPLE CORRELATION COEFFICIENT
     1SQUARED =',F6.2,' PERCENT')
C
C     COMPUTE THE F VALUE OF THE INDEPENDENT TERM IN REGRESSION THAT HAS
C         THE LEAST CONTRIBUTION TO THE REDUCTION OF VARIANCE
C
      FTEST2=ABS(VMIN)*DF/A(N,N)
C
C     TEST THIS COMPUTED F VALUE AGAINST A SPECIFIED TABULAR F VALUE -
C         IF THE COMPUTED VALUE IS LESS THAN THE TABULAR VALUE, THEN
C         REMOVE THE INDEPENDENT TERM FROM REGRESSION - OTHERWISE GO
C         TEST FOR THE ADDITION OF ANOTHER INDEPENDENT TERM
C
      IF(FTEST2.GE.F2)GO TO 480
C
C     PRINT THE STEP NUMBER, THE INDEPENDENT TERM REMOVED FROM
C         REGRESSION, AND THE COMPUTED F VALUE OF THE TERM
C
      WRITE(3,3300)STEPNO,MIN,FTEST2
 3300 FORMAT(///1X,'STEP NO. ',I3/3X,'VARIABLE REMOVED',I4/
     13X,'F LEVEL',F8.4)
C
C     SET THE SUBSCRIPT OF THE PIVOT ELEMENT EQUAL TO THE NUMBER OF THE
C         TERM JUST REMOVED FROM REGRESSION
C
```

```
      K=MIN
C
C     INCREASE THE RESIDUAL DEGREES OF FREEDOM BY ONE TO ACCOUNT FOR THE
C         TERM JUST REMOVED FROM REGRESSION
C
      DF=DF+1
C
C     GO DO A TRANSFORMATION ON MATRIX A
C
      GO TO 560
C
C     COMPUTE THE F VALUE OF THE MOST HIGHLY CORRELATED TERM THAT IS NOT
C         ALREADY IN REGRESSION
C
  480 FTEST1=VMAX*(DF-1)/(A(N,N)-VMAX)
C
C     TEST WHETHER THE TERM SHOULD ENTER THE REGRESSION EQUATION BY
C         CHECKING ITS COMPUTED F VALUE AGAINST A SPECIFIED TABULAR F
C         VALUE - IF THE COMPUTED VALUE IS GREATER THAN THE TABULAR
C         VALUE, THEN GO ENTER THE TERM INTO REGRESSION - OTHERWISE
C         REPOSITION THE DATA POINTER ON THE UNIT SPECIFIED TO THE
C         LOCATION OF THE FIRST INPUT DATA ELEMENT
C
      IF(FTEST1.GT.F1) GO TO 550
      REWIND 4
C
C     CHECK WHETHER THE RESIDUALS SHOULD BE PRINTED - IF NOT, THEN GO
C         READ IN THE NEXT CONTROL CARD
C
      IF(IRESID.EQ.0) GO TO 10
C
```

```
C     PRINT THE HEADING FOR THE RESIDUALS
C
      WRITE(3,3310)
 3310 FORMAT(3X,'PREDICTED VS. ACTUAL RESULTS'/10X,'OBSERVATION',10X,
     1'ACTUAL',11X,'PREDICTED',11X,'DEVIATION')
      DO 540 IK=1,M
C
C     READ AN OBSERVATION OF INPUT DATA FROM THE UNIT SPECIFIED
C
      READ(4,FMT)(Z(J),J=1,N1)
C
C     MOVE THE VALUE OF THE DEPENDENT VARIABLE TO A TEMPORARY STORAGE
C         AREA
C
      TEMP=Z(DEPVAR)
C
C     MOVE THE VALUE OF EACH INPUT VARIABLE POSITIONED BEYOND THE
C         DEPENDENT VARIABLE TO THE LOCATION OF THE INPUT VARIABLE
C         THAT PRECEDED IT
C
      DO 485 IL=DEPVAR,N1M1
  485 Z(IL)=Z(IL+1)
C
C     MOVE THE VALUE OF THE DEPENDENT VARIABLE TO THE LOCATION OF THE
C         LAST INPUT VARIABLE
C
      Z(N1)=TEMP
C
C     INITIALIZE SUMPR ( THE SUM OF INDEPENDENT TERMS TIMES THEIR
C         RESPECTIVE REGRESSION COEFFICIENTS FOR THE OBSERVATION BEING
C         CONSIDERED )
```

```
C
      SUMPR=0.0
      DO 530 J=1,INEQ
C
C     IF NORMAL MODE IS USED SET IRET EQUAL TO THE TERM NUMBER BEING
C         CONSIDERED - OTHERWISE SET IRET EQUAL TO THE NUMBER THE TERM
C         WOULD HAVE HAD IF THE MODE WAS NORMAL
C
      IF(MODE.EQ.INORM) GO TO 490
      IRET=LIST(INVAR(J))
      GO TO 500
  490 IRET=INVAR(J)
C
C     DETERMINE THE POWERS OF THE INDEPENDENT VARIABLES THAT CORRESPOND
C         TO THE TERM BEING CONSIDERED - THIS IS DONE BY REPEATED
C         DIVISIONS
C
  500 DO 510 IJ=1,N1M1
C
C     DIVIDE BY (MAXIMUM POWER OF THE PARTICULAR INDEPENDENT VARIABLE +
C         ONE) AND THEN TRUNCATE
C
      IZ=IRET/(MAXP(N1-IJ)+1)
C
C     SUBTRACT THE TRUNCATED VALUE FROM THE ORIGINAL VALUE TO OBTAIN
C         THE PARTICULAR POWER
C
      P(N1-IJ)=IRET-IZ*(MAXP(N1-IJ)+1)
C
C     DROP THE PART OF THE TERM ALREADY DIVIDED OUT
C
```

```
  510 IRET=IZ
C
C     INITIALIZE PROD (THE PRODUCT OF THE INDEPENDENT VARIABLES TIMES
C         THEIR RESPECTIVE POWERS FOR THE INDEPENDENT TERM BEING
C         CALCULATED)
C
      PROD=1.0
      DO 520 II=1,N1M1
C
C     IF THE POWER OF THE INDEPENDENT VARIABLE IN QUESTION IS ZERO, THEN
C         IT SHOULD NOT BE CONSIDERED WHEN BUILDING AN INDEPENDENT TERM
C
      IF(P(II).EQ.0) GO TO 520
C
C     UPDATE PROD BY MULTIPLYING IT BY THE INDEPENDENT VARIABLE BEING
C         CONSIDERED TIMES ITS CORRESPONDING POWER
C
      PROD=PROD*Z(II)**P(II)
  520 CONTINUE
C
C     UPDATE SUMPR BY ADDING IN PROD TIMES THE REGRESSION COEFFICIENT OF
C         THE INDEPENDENT TERM BEING CONSIDERED
C
  530 SUMPR=SUMPR+PROD*B(J)
C
C     CALCULATE THE PREDICTED VALUE OF THE DEPENDENT VARIABLE
C
      YHAT=BZERO+SUMPR
C
C     CALCULATE THE DEVIATION BETWEEN THE ACTUAL VALUE AND THE PREDICTED
C         VALUE OF THE DEPENDENT VARIABLE FOR THE OBSERVATION BEING
```

```
C         CONSIDERED
C
      DEV=Z(N1)-YHAT
C
C     PRINT THE OBSERVATION NUMBER, THE ACTUAL VALUE, THE PREDICTED
C         VALUE, AND THE DEVIATION BETWEEN THE ACTUAL AND PREDICTED
C         VALUE OF THE DEPENDENT VARIABLE
C
  540 WRITE(3,3320)IK,Z(N1),YHAT,DEV
 3320 FORMAT(12X,I5,4X,F16.5,4X,F16.5,4X,F16.5)
C
C     REPOSITION THE DATA COUNTER TO THE LOCATION OF THE FIRST INPUT
C         DATA ELEMENT ON THE UNIT SPECIFIED
C
      REWIND 4
C
C     GO READ THE NEXT CONTROL CARD
C
      GO TO 10
C
C     PRINT THE STEP NUMBER, TERM ENTERING THE REGRESSION EQUATION, AND
C         THE COMPUTED F VALUE
C
  550 WRITE(3,3330)STEPNO,MAX,FTEST1
 3330 FORMAT(///1X,'STEP NO.',I3/3X,'VARIABLE ENTERING',I4/
     13X,'F LEVEL',F8.4)
C
C     SET THE SUBSCRIPT OF THE PIVOT ELEMENT EQUAL TO THE NUMBER OF THE
C         TERM ENTERING REGRESSION
C
      K=MAX
```

```
C
C     DECREASE THE RESIDUAL DEGREES OF FREEDOM BY ONE TO ACCOUNT FOR THE
C         TERM ENTERING REGRESSION
C
      DF=DF-1
C
C     SET THE PIVOT ELEMENT (PIVOT) FOR TRANSFORMATION OF MATRIX A
C         (THE FIRST TIME MATRIX A = MATRIX R)
C
  560 PIVOT=A(K,K)
      DO 590 I=1,N
      IF(I.EQ.K)GO TO 590
C
C     COMPUTE THE ELEMENT NEEDED IN THE PIVOT COLUMN FOR THE ROW OF
C         MATRIX A BEING PROCESSED
C
      PIVCOL=A(I,K)
      DO 580 J=1,N
      IF(J.EQ.K)GO TO 570
C
C     CALCULATE NEW A(I,J) FOR I NOT EQUAL TO K AND J NOT EQUAL TO K
C
      A(I,J)=(A(I,J)*PIVOT-PIVCOL*A(K,J))/PIVOT
      GO TO 580
C
C     CALCULATE NEW A(I,J) FOR I NOT EQUAL K AND J EQUAL K
C
  570 A(I,J)=-A(I,K)/PIVOT
  580 CONTINUE
  590 CONTINUE
      DO 610 J=1,N
```

```
      IF(J.EQ.K) GO TO 600
C
C     CALCULATE NEW A(I,J) FOR I EQUAL K AND J NOT EQUAL K
C
      A(K,J)=A(K,J)/PIVOT
      GO TO 610
C
C     CALCULATE NEW A(K,K)
C
  600 A(K,K)=1.0/PIVOT
  610 CONTINUE
      GO TO 380
  620 CALL EXIT
      END
```

```
      SUBROUTINE GENLST(MODE, N1M1, N, NM1, NCARD, MAXP, LIST, ELM, ALPHA)
C
C     THIS SUBROUTINE GENERATES A LIST OF TERM NUMBERS - WHEN IN
C         THE NORMAL MODE MANY COMBINATIONS OF THE POWERS OF THE
C         INDEPENDENT VARIABLES ARE POSSIBLE WITHOUT EXCEEDING THE
C         MAXIMUM POWER OF ANY INDEPENDENT VARIABLE - EACH POSSIBLE
C         COMBINATION REPRESENTS ONE INDEPENDENT TERM - ALL POSSIBLE
C         COMBINATIONS MAKE UP A COMPLETE SET OF TERMS - THESE
C         INDEPENDENT TERMS ARE GENERATED IN AN ORDERLY FASHION SO THAT
C         EACH INDEPENDENT TERM IS GIVEN A CONSECUTIVE NUMBER - WHEN IN
C         ACCEPT OR DELETE MODE, ONLY A SUBSET OF THE COMPLETE SET OF
C         TERMS IS DESIRED - THUS THE LIST IS COMPOSED OF THE NUMBER OF
C         THE TERMS IN A COMPLETE SET THAT CORRESPOND TO THE DESIRED
C         TERMS
C
C     ALPHA(I)=ARRAY OF ALPHA INFORMATION THAT INDICATES THE INDEPENDENT
C         TERMS THAT SHOULD BE DELETED OR ACCEPTED
C     ELM(I, J)=ARRAY THAT CONTAINS THE VALUES OF THE POWERS OF THE
C         INDEPENDENT TERMS THAT ARE TO BE DELETED OR ACCEPTED
C     I=TERM NUMBER AND COUNTER
C     ICOL=VALUE OF THE COLUMN OF ARRAY ELM CURRENTLY BEING PROCESSED
C     II =DO LOOP PARAMETER
C     IJ =DO LOOP PARAMETER
C     IROW=VALUE OF THE ROW OF ARRAY ELM CURRENTLY BEING PROCESSED
C     ITEST=INDEX THAT DENOTES THE INDEPENDENT VARIABLE PRESENTLY BEING
C         PROCESSED
C     K=NUMBER OF CHARACTERS OF ALPHA INFORMATION READ FROM THE VARIABLE
C         MODE CONTROL DATA CARDS
C     LIST(I)=ARRAY THAT CONTAINS THE VALUES OF THE INDEXES FOR THOSE
C         INDEPENDENT TERMS THAT ARE TO BE CONSIDERED IN THE REGRESSION
C         MODEL
```

```
C     LL=SUBSCRIPT USED FOR THE VECTOR LIST - ALSO COUNTS THE NUMBER OF
C         INDEPENDENT TERMS
C     N=NUMBER OF TERMS
C     NM1=NUMBER OF INDEPENDENT TERMS
C     P(I)=POWERS OF THE INDEPENDENT VARIABLES NEEDED TO COMPOSE THE
C         INDEPENDENT TERM BEING CONSIDERED
C
      DATA IBLK/4H    /,ICOM/4H,   /,IAST/4H*   /,ISHIFT/Z01000000/,
     1IACCE/4HACCE/,IDELE/4HDELE/
      INTEGER P(15),MAXP(15),ALPHA(3000),ELM(200,15),LIST(150)
C
C     DETERMINE THE NUMBER OF CHARACTERS OF ALPHA INFORMATION ON THE
C         VARIABLE MODE CONTROL DATA CARDS WHICH SPECIFY THE
C         INDEPENDENT TERMS TO BE DELETED OR ACCEPTED
C
      K=NCARD*80
C
C     PLACE ALL CHARACTERS FROM THE VARIABLE MODE CONTROL DATA CARDS IN
C         THE ARRAY ALPHA
C
      READ(1,1000)(ALPHA(J),J=1,K)
 1000 FORMAT(80A1)
C
C     INITIALIZE THE ROW AND COLUMN OF THE ARRAY ELM
C
      IROW=1
      ICOL=0
      DO 40 I=1,K
C
C     IF THE CHARACTER BEING TESTED IS A BLANK, THEN GO TO THE NEXT
C         CHARACTER
```

```
C
      IF(ALPHA(I).EQ.IBLK) GO TO 40
C
C     IF THE CHARACTER BEING TESTED IS A COMMA, THEN THE FIRST NUMERIC
C         CHARACTER TO FOLLOW WILL BE A POWER OF THE FIRST INDEPENDENT
C         VARIABLE OF THE NEXT INDEPENDENT TERM TO BE ACTED UPON -
C         OTHERWISE THE CHARACTER DENOTES A POWER OF A INDEPENDENT
C         VARIABLE FOR THE TERM BEING CONSIDERED
C
      IF(ALPHA(I).NE.ICOM) GO TO 20
C
C     IF THE NUMBER OF CHARACTERS DETECTED BETWEEN THE TWO
C         PREVIOUS COMMAS IS NOT EQUAL TO THE NUMBER OF INDEPENDENT
C         VARIABLES, THEN PRINT AN ERROR MESSAGE AND TERMINATE
C
      IF(ICOL.EQ.N1M1) GO TO 10
      WRITE(3,3000)
 3000 FORMAT(' INCORRECT NUMBER OF CHARACTERS BETWEEN COMMAS ON A VARIAB
     1LE MODE CARD')
      STOP
C
C     SINCE THE LAST CHARACTER TESTED WAS A COMMA, INCREMENT THE ROW
C         INDEX AND RESET THE COLUMN INDEX OF ARRAY ELM FOR THE NEXT
C         TERM - THEN GO TO TEST THE NEXT CHARACTER
C
   10 IROW=IROW+1
      ICOL=0
      GO TO 40
C
C     SINCE THE LAST CHARACTER TESTED WAS A POWER OF AN INDEPENDENT
C         VARIABLE, INCREMENT OVER ONE COLUMN IN ARRAY ELM IN ORDER TO
```

```
C         PLACE IN IT THE VALUE OF THE POWER
C
   20 ICOL=ICOL+1
C
C     THE ASTERISK (*) IS USED TO DENOTE A VARIABLE POWER THAT WILL
C         ALWAYS BE ACKNOWLEDGED TO MEAN ANY POWER OF THE INDEPENDENT
C         VARIABLE IN QUESTION
C
C     IF THE POWER DETECTED IS NOT AN ASTERISK, THEN CONVERT THE POWER
C         FROM ALPHA INFORMATION TO A NUMERIC NUMBER AND PLACE THE
C         VALUE IN ARRAY ELM, OTHERWISE PUT AN ASTERISK IN THE
C         PARTICULAR ELEMENT OF ARRAY ELM
C
      IF(ALPHA(I).EQ.IAST) GO TO 30
C
C     SHIFT ALPHA(I) FROM THE HIGH ORDER BYTE OF THE WORD TO THE LOW
C         ORDER BYTE OF THE WORD
C
      ELM(IROW,ICOL)=15-IABS(ALPHA(I)/ISHIFT)
      GO TO 40
   30 ELM(IROW,ICOL)=IAST
   40 CONTINUE
C
C     INITIALIZE THE POWER OF ALL INDEPENDENT VARIABLES TO ZERO - THEN
C         SET THE POWER OF THE LAST INDEPENDENT VARIABLE TO ONE SO
C         THAT THE FIRST INDEPENDENT TERM MAY BE REPRESENTED
C
      DO 50 I=1,N1M1
   50 P(I)=0
      P(N1M1)=1
C
```

```
C     INITIALIZE THE TERM COUNTER (I) AND THE INDEX USED TO INCREMENT
C         THROUGH ARRAY LIST (LL)
C
      I=0
      LL=0
C
C     INITIALIZE THE INDEX THAT DENOTES THE INDEPENDENT VARIABLE
C         PRESENTLY BEING PROCESSED
C
   60 ITEST=N1M1
C
C     INCREMENT THE TERM COUNTER
C
   70 I=I+1
C
C     CHECK WHETHER THE INDEPENDENT TERM BEING GENERATED IS REPRESENTED
C         IN ARRAY ELM - II IS THE ROW INDEX AND IJ IS THE COLUMN INDEX
C         FOR THE ARRAY ELM
C
      DO 90 II=1,IROW
      DO 80 IJ=1,ICOL
C
C     IF AN ASTERISK IS ENCOUNTERED IN AN ELEMENT OF ARRAY ELM, THEN
C         THIS IS CONSIDERED A MATCH FOR THE POWER OF THE INDEPENDENT
C         VARIABLE BEING CONSIDERED - PROCEED TO CHECK THE NEXT
C         INDEPENDENT VARIABLE
C
      IF(ELM(II,IJ).EQ.IAST)GO TO 80
C
C     IF THE VALUE OF THE POWER OF THE INDEPENDENT VARIABLE BEING
C         CONSIDERED IS EQUAL TO THE VALUE OF THE ELEMENT OF ARRAY ELM
```

```
C           BEING CONSIDERED, THEN A MATCH OCCURS AND PROCEED TO CHECK
C           THE POWER OF THE NEXT INDEPENDENT VARIABLE - OTHERWISE SKIP
C           TO THE NEXT ROW OF ARRAY ELM AND START CHECKING IT
C
      IF(ELM(II,IJ).NE.P(IJ)) GO TO 90
   80 CONTINUE
C
C     NOW THE POWERS OF ALL INDEPENDENT VARIABLES HAVE MATCHED AND THE
C           TERM BEING CONSIDERED IS REPRESENTED IN ARRAY ELM
C
      GO TO 100
   90 CONTINUE
C
C     THE INDEPENDENT TERM BEING CONSIDERED WAS NOT REPRESENTED BY ANY
C           ROW OF THE ARRAY ELM
C
C     DETERMINE WHETHER THE VARIABLE MODE IS ACCEPT OR DELETE
C
      IF(MODE.EQ.IACCE) GO TO 120
      GO TO 110
C
C     DETERMINE WHETHER THE VARIABLE MODE IS ACCEPT OR DELETE
C
  100 IF(MODE.EQ.IDELE) GO TO 120
C
C     INCREMENT THE INDEX USED FOR THE VECTOR LIST
C
  110 LL=LL+1
C
C     PUT THE INDEPENDENT TERM IN THE LIST - THE TERM NUMBER IS PUT IN
C           THE LIST IF IT WAS REPRESENTED IN A ROW OF ARRAY ELM AND WAS
```

```
C         IN ACCEPT MODE, OR IF IT WAS NOT REPRESENTED IN A ROW OF
C         ARRAY ELM AND WAS IN DELETE MODE
C
      LIST(LL)=I
C
C     PROCEED TO CHANGE THE POWERS OF THE INDEPENDENT VARIABLES SO THAT
C         THE NEXT INDEPENDENT TERM CAN BE REPRESENTED
C
C     CHECK WHETHER THE POWER LAST INCREMENTED HAS REACHED THE MAXIMUM
C         POWER FOR THAT INDEPENDENT VARIABLE
C
  120 IF(P(ITEST).EQ.MAXP(ITEST)) GO TO 130
C
C     AGAIN INCREMENT THE POWER OF THE SAME INDEPENDENT VARIABLE WHOSE
C         POWER WAS LAST INCREMENTED
C
      P(ITEST)=P(ITEST)+1
      GO TO 70
C
C     TEST WHETHER THE POWER OF THE PRECEDING INDEPENDENT VARIABLE HAS
C         REACHED ITS MAXIMUM POWER
C
  130 IF(P(ITEST-1).NE.MAXP(ITEST-1)) GO TO 140
C
C     TEST WHETHER THE FIRST INDEPENDENT VARIABLE HAS REACHED ITS
C         MAXIMUM POWER - ALL INDEPENDENT TERMS HAVE BEEN CONSIDERED
C         WHEN THIS CONDITION OCCURS
C
      IF(ITEST.EQ.2) GO TO 160
C
C     DECREMENT THE COUNTER TO THE PRECEDING VARIABLE
```

```
C
      ITEST=ITEST-1
      GO TO 130
C
C     REINITIALIZE ALL INDEPENDENT VARIABLES FROM THE LAST ONE
C         CONSIDERED UP TO THE LAST ONE
C
  140 DO 150 II=ITEST, N1M1
  150 P(II)=0
C
C     INCREASE THE POWER OF THE PRECEDING INDEPENDENT VARIABLE BY ONE
C
      P(ITEST-1)=P(ITEST-1)+1
      GO TO 60
C
C     COMPUTE THE NUMBER OF INDEPENDENT TERMS
C
  160 NM1=LL
C
C     COMPUTE THE NUMBER OF TERMS
C
      N=LL+1
      RETURN
      END
```

```
      SUBROUTINE GENTRM(N1M1, NM1, MAXP, LIST, Z, X)
C
C     THIS SUBROUTINE IS USED ONLY WHEN IN ACCEPT OR DELETE MODE - ALL
C         INDEPENDENT TERMS AND THE DEPENDENT TERM ARE COMPUTED
C
C     I=DO LOOP PARAMETER
C     II=DO LOOP PARAMETER
C     IJ=DO LOOP PARAMETER
C     IRET=NUMBER OF THE TERM IN THE NORMAL MODE THAT IS TO
C         BE RETRIEVED AS POWERS OF THE INDEPENDENT VARIABLES
C     IZ=FRACTIONAL PART OF IRET TO STILL BE RETRIEVED
C     N1=NUMBER OF INPUT VARIABLES
C     P(I)=POWERS OF THE INDEPENDENT VARIABLES NEEDED TO COMPOSE THE
C         INDEPENDENT TERM BEING CONSIDERED
C     X(I)=TERM I OF A GIVEN OBSERVATION
C
      DIMENSION Z(15), X(150)
      INTEGER P(15), MAXP(15), LIST(150)
C
C     COMPUTE THE NUMBER OF INDEPENDENT VARIABLES
C
      N1=N1M1+1
C
C     I REPRESENTS THE INDEPENDENT TERM BEING CONSIDERED
C
      DO 20 I=1, NM1
C
C     SET IRET EQUAL TO THE TERM NUMBER THAT WOULD HAVE BEEN GENERATED
C         IF THE INDEPENDENT TERM BEING CONSIDERED WAS IN THE NORMAL
C         MODE
C
```

```
      IRET=LIST(I)
C
C     THIS DO LOOP CALCULATES THE POWERS OF THE INDEPENDENT VARIABLE FOR
C         THE INDEPENDENT TERM BEING CONSIDERED
C
      DO 10 IJ=1,N1M1
C
C     DIVIDE BY (MAXIMUM POWER + 1) AND THEN TRUNCATE
C
      IZ=IRET/(MAXP(N1-IJ)+1)
C
C     SUBTRACT OFF THE TRUNCATED VALUE TO OBTAIN THE POWER OF THE
C         INDEPENDENT VARIABLE IN QUESTION
C
      P(N1-IJ)=IRET-IZ*(MAXP(N1-IJ)+1)
C
C     DROP THE PART OF THE TERM ALREADY DIVIDED OUT
C
   10 IRET=IZ
C
C     INITIALIZE THE VALUE OF THE INDEPENDENT TERM UNDER CONSIDERATION
C
      X(I)=1.0
      DO 20 II=1,N1M1
C
C     IF THE POWER OF THE INDEPENDENT VARIABLE IN QUESTION IS ZERO, THEN
C         IT SHOULD NOT BE CONSIDERED WHEN BUILDING AN INDEPENDENT TERM
C
      IF(P(II).EQ.0) GO TO 20
C
C     UPDATE THE VALUE OF THE INDEPENDENT TERM UNDER CONSIDERATION
```

```
C
      X(I)=X(I)*Z(II)**P(II)
   20 CONTINUE
C
C     COMPUTE THE DEPENDENT TERM
C
      X(NM1+1)=Z(N1)
      RETURN
      END
```

```
      SUBROUTINE EXPAND(N1M1,MAXP,Z,X)
C
C     THIS SUBROUTINE IS USED ONLY IN NORMAL MODE - IT COMPUTES THE
C         VALUE OF THE COMPLETE SET OF INDEPENDENT TERMS BY EXPANDING
C         TO ALL POSSIBLE COMBINATIONS OF ALL POWERS OF THE INDEPENDENT
C         VARIABLES UP TO THEIR MAXIMUM POWERS - THEN THE DEPENDENT
C         TERM IS COMPUTED
C
C     I=TERM NUMBER AND COUNTER
C     II=DO LOOP PARAMETER
C     ITEST=INDEX THAT DENOTES THE INDEPENDENT VARIABLE PRESENTLY BEING
C         PROCESSED
C     P(I)=POWERS OF THE INDEPENDENT VARIABLES NEEDED TO COMPOSE THE
C         INDEPENDENT TERM BEING CONSIDERED
C     X(I)=TERM I OF A GIVEN OBSERVATION
C
      DIMENSION Z(15),X(125)
      INTEGER P(15),MAXP(15)
C
C     INITIALIZE THE POWER OF ALL INDEPENDENT VARIABLES TO ZERO
C
      DO 10 I=1,N1M1
   10 P(I)=0
C
C     SET THE POWER OF THE LAST INDEPENDENT VARIABLE TO ONE, THUS
C         REPRESENTING THE FIRST INDEPENDENT TERM
C
      P(N1M1)=1
C
C     INITIALIZE THE TERM COUNTER
C
```

```
      I=0
C
C     INITIALIZE THE INDEX THAT DENOTES THE INDEPENDENT VARIABLE
C         PRESENTLY BEING PROCESSED
C
   20 ITEST=N1M1
C
C     INCREMENT THE TERM COUNTER
C
   30 I=I+1
C
C     INITIALIZE THE VALUE OF THE INDEPENDENT TERM UNDER CONSIDERATION
C
      X(I)=1.0
C
C     II IS THE INDEX THAT DENOTES THE INDEPENDENT VARIABLE BEING
C         PROCESSED
C
      DO 40 II=1,N1M1
C
C     IF THE POWER OF THE INDEPENDENT VARIABLE IN QUESTION IS ZERO, THEN
C         IT SHOULD NOT BE CONSIDERED WHEN BUILDING AN INDEPENDENT TERM
C
      IF(P(II).EQ.0) GO TO 40
C
C     UPDATE THE VALUE OF THE INDEPENDENT TERM BEING CALCULATED BY
C         MULTIPLYING BY AN INDEPENDENT VARIABLE RAISED TO A PARTICULAR
C         POWER
C
      X(I)=X(I)*Z(II)**P(II)
   40 CONTINUE
```

```
C
C     PROCEED TO CHANGE THE POWERS OF THE INDEPENDENT VARIABLES SO THAT
C         THE NEXT INDEPENDENT TERM CAN BE REPRESENTED
C
C     CHECK WHETHER THE POWER OF THE INDEPENDENT VARIABLE LAST
C         INCREMENTED HAS REACHED THE MAXIMUM POWER FOR THAT
C         INDEPENDENT VARIABLE
C
      IF(P(ITEST).EQ.MAXP(ITEST)) GO TO 50
C
C     AGAIN INCREMENT THE POWER OF THE SAME INDEPENDENT VARIABLE WHOSE
C         POWER WAS LAST INCREMENTED
C
      P(ITEST)=P(ITEST)+1
      GO TO 30
C
C     TEST WHETHER THE POWER OF THE PRECEDING INDEPENDENT VARIABLE HAS
C         REACHED ITS MAXIMUM POWER
C
   50 IF(P(ITEST-1).NE.MAXP(ITEST-1)) GO TO 60
C
C     TEST WHETHER THE FIRST INDEPENDENT VARIABLE HAS REACHED ITS
C         MAXIMUM POWER - ALL INDEPENDENT TERMS OF THE COMPLETE SET
C         HAVE BEEN CONSIDERED WHEN THIS CONDITION OCCURS
C
      IF(ITEST.EQ.2) GO TO 80
C
C     DECREMENT THE COUNTER TO THE PRECEDING INDEPENDENT VARIABLE
C
      ITEST=ITEST-1
      GO TO 50
```

```
C
C     REINITIALIZE ALL INDEPENDENT VARIABLES FROM THE LAST ONE
C           CONSIDERED UP TO THE LAST ONE
C
   60 DO 70 II=ITEST, N1M1
   70 P(II)=0
C
C     INCREASE THE POWER OF THE PRECEDING INDEPENDENT VARIABLE BY ONE
C           AND GO PROCESS THE NEXT INDEPENDENT TERM
C
      P(ITEST-1)=P(ITEST-1)+1
      GO TO 20
C
C     COMPUTE THE DEPENDENT TERM
C
   80 X(I+1)=Z(N1M1+1)
      RETURN
      END
```